Basic BioGrid

Rick Stevens
Argonne National Laboratory
University of Chicago

Outline

- Biology: The Science for the 21st Century
- Sizing up the opportunities for BioGrid
- Biology is a different way of thinking
- Systems biology and whole cell modeling
- Requirements for the BioGrid
- A modest proposal
- Some recommended reading
- Conclusions

Many BioGrid Projects

biogrid Living Energy Everyday

- EUROGRID BioGRID BIOGRID
- Asia Pacific BioGRID
- NC BioGrid
- Bioinformatics Research Network BIRN
- Osaka University Biogrid
- Indiana University BioArchive BioGrid

Biogrid
バイオグリッド研究会

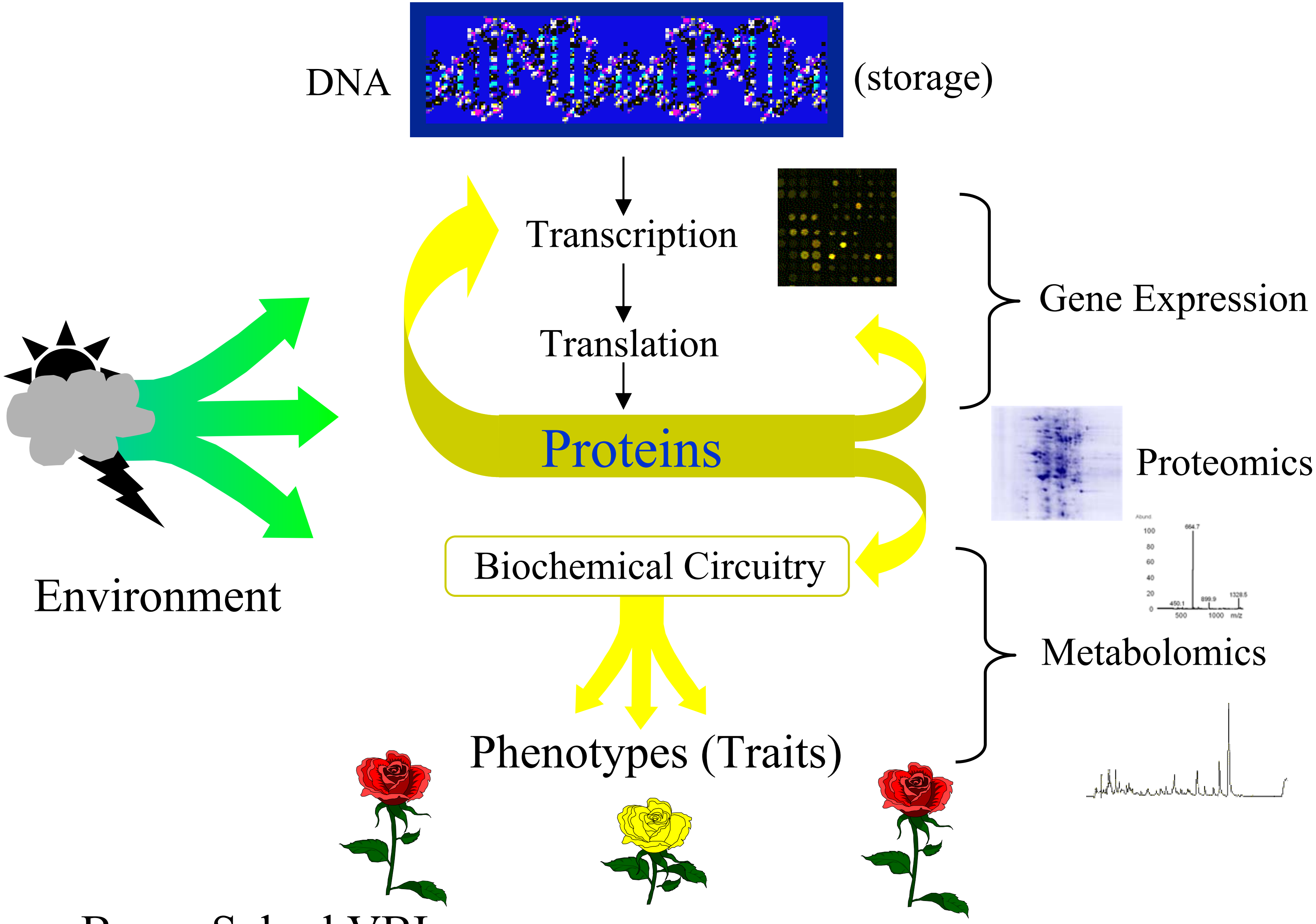
Take The World of New Biology In Your Hand
GENOMICS • PROTEOMICS • PHARMACOGENOMICS • LABORATORY INFRASTRUCTURE

The New Biology

- Genomics
- Functional Genomics
- Proteomics
- Structural Biology
- Gene Expression
- Metabolomics
- Advanced Imaging

- High-throughput methods
  - Low cost
  - Robotics
- Bioinformatics driven
- Quantitative
- Enables a systems view
- Basis for integrative understanding
  - Global state
  - Time dependent

Predicting Life Processes: Reverse Engineering Living Systems



DNA (storage)

Transcription

Translation

Proteins

Biochemical Circuitry

Phenotypes (Traits)

Environment

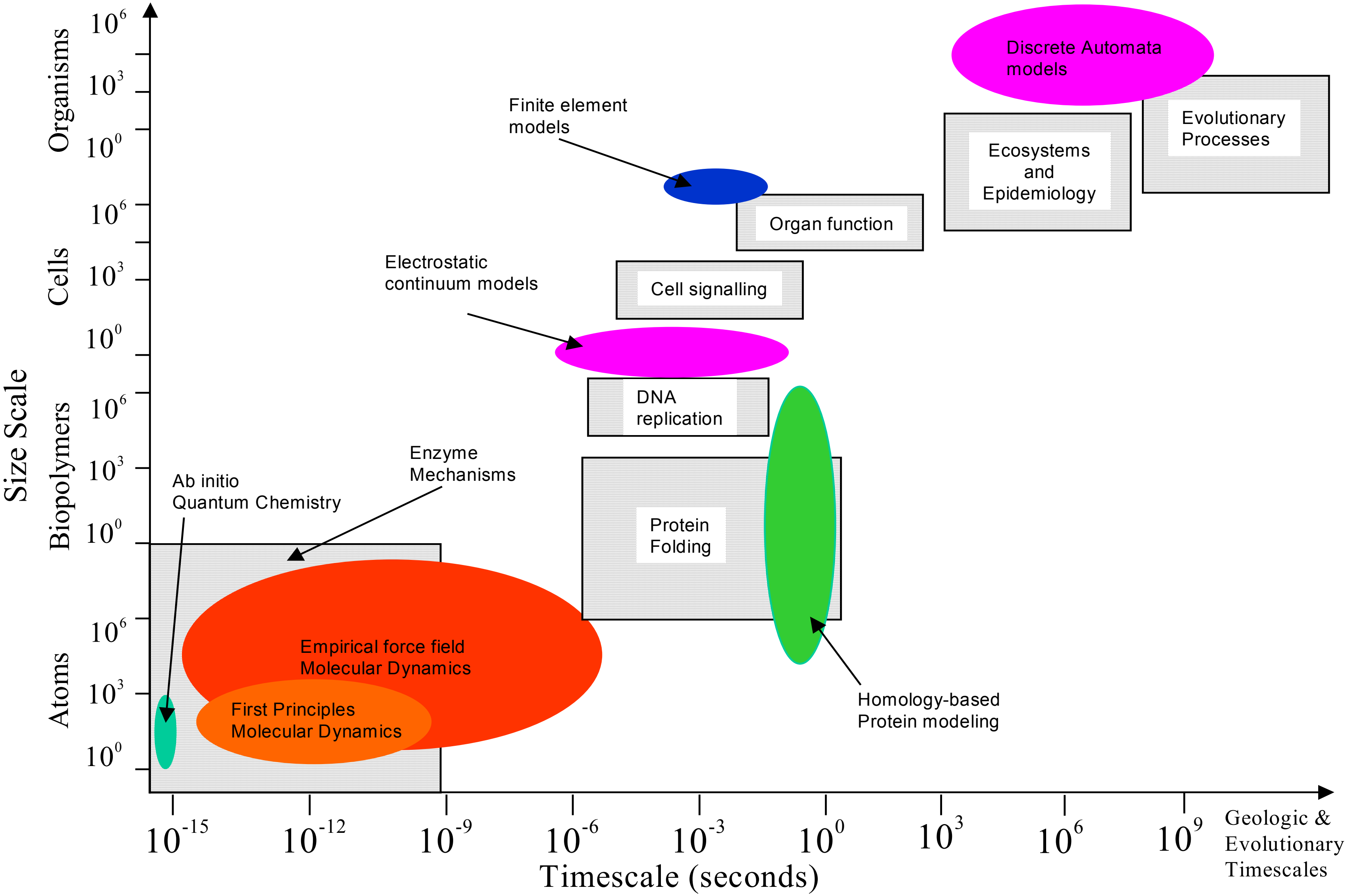
Gene Expression

Proteomics

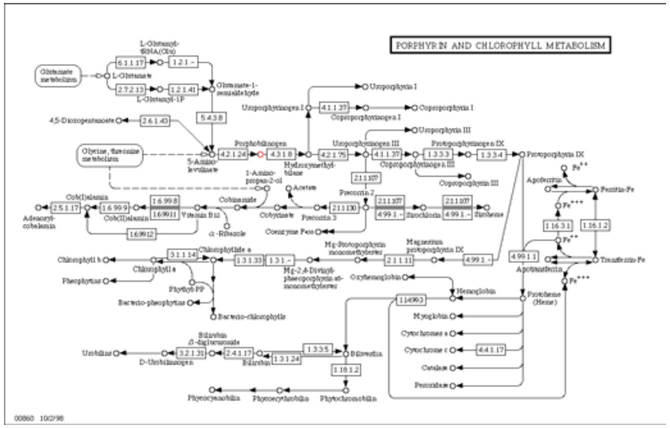
Metabolomics

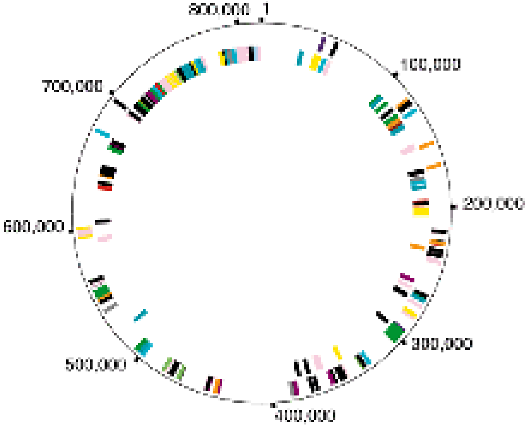
From Bruno Sobral VBI

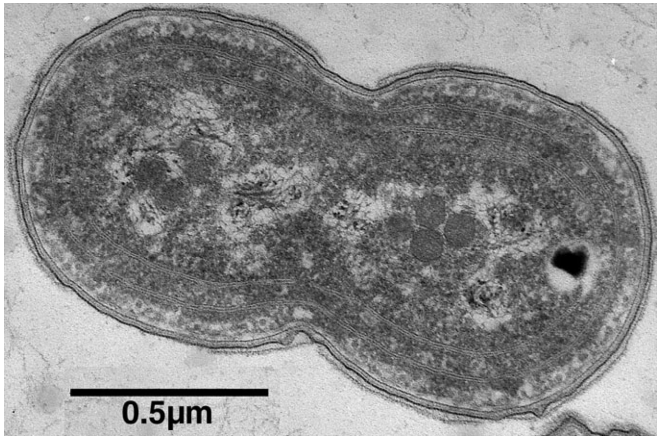
24 Orders Magnitude of Spatial and Temporal Range



Genes → Cell Networks → Organisms → Populations → Ecosystems







Rick Stevens

A

Computational analysis and simulation have important roles in the study of each step in the hierarchy of biological function

DNA sequence

Sequence Annotation

Protein sequence and regulation

Promoter

T
A
T
A

Message

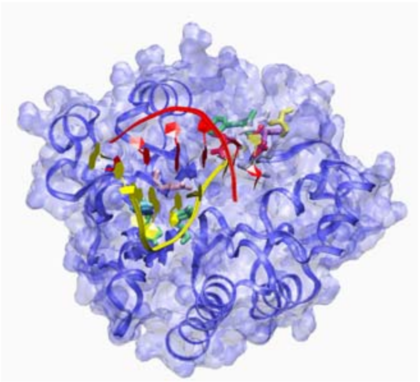
C A G Q
T A C Y
C G T R

Homology based protein structure prediction

Protein structure

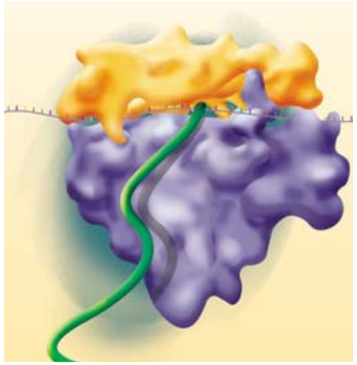
Molecular simulations

Protein/enzyme function



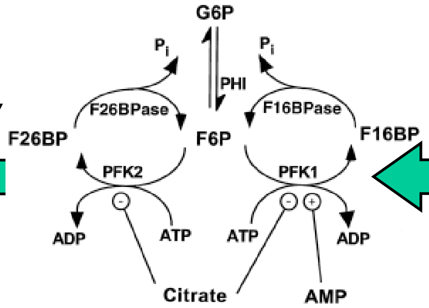
Expt. data integration

Multi-protein machines



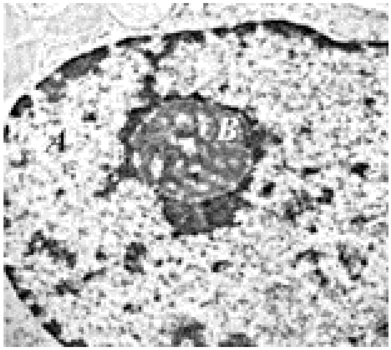
Network analysis

Metabolic pathways & regulatory networks



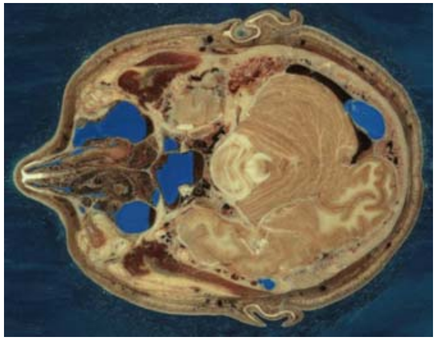
Pathway simulations

Bacteria and cells



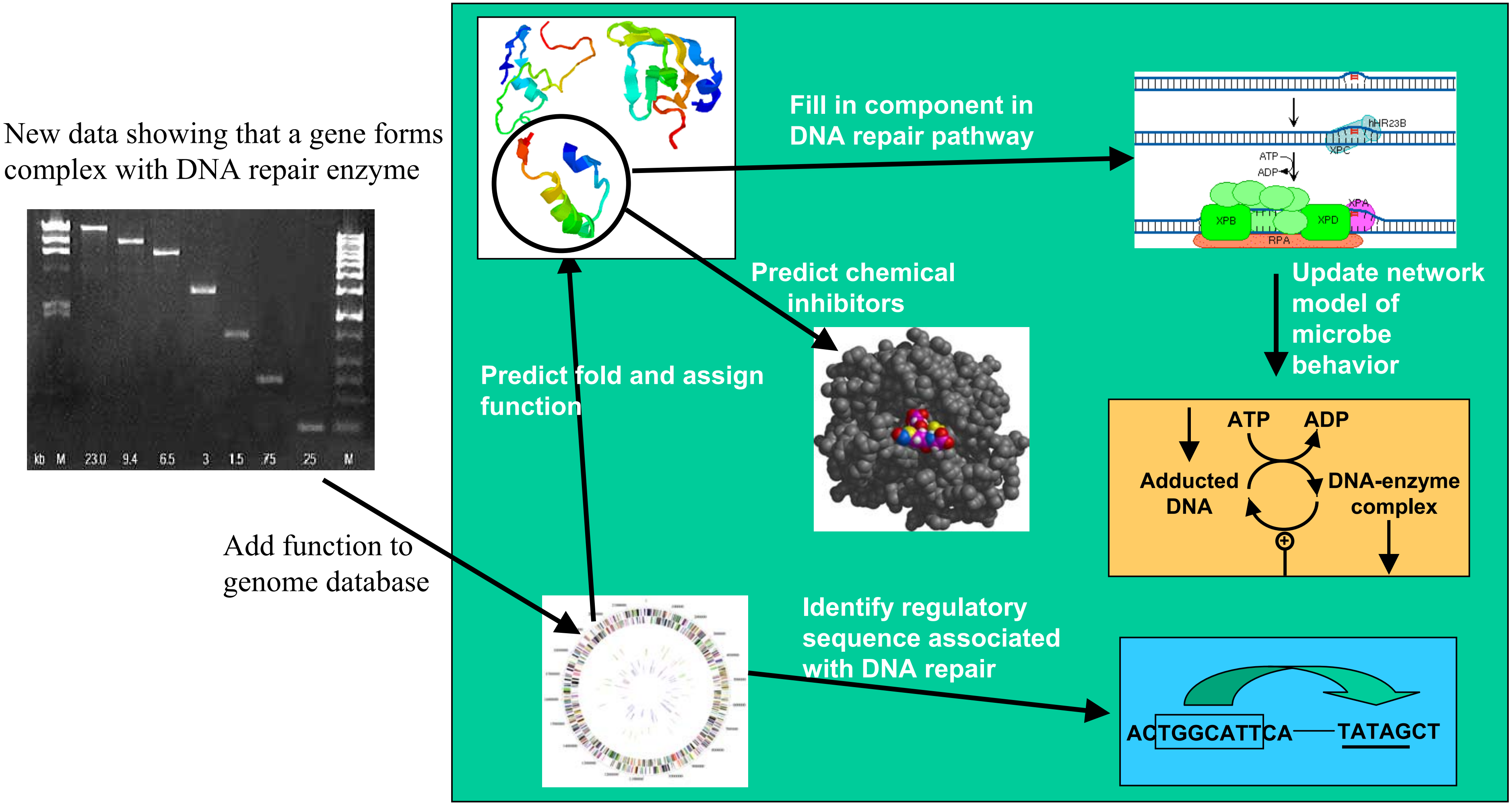
Organism simulations

Bacterial communities & multicellular organisms



The New Biology in Action

Hypothetical example:



New data showing that a gene forms complex with DNA repair enzyme

Add function to genome database

Predict fold and assign function

Fill in component in DNA repair pathway

Predict chemical inhibitors

Update network model of microbe behavior

Identify regulatory sequence associated with DNA repair

ATP ADP

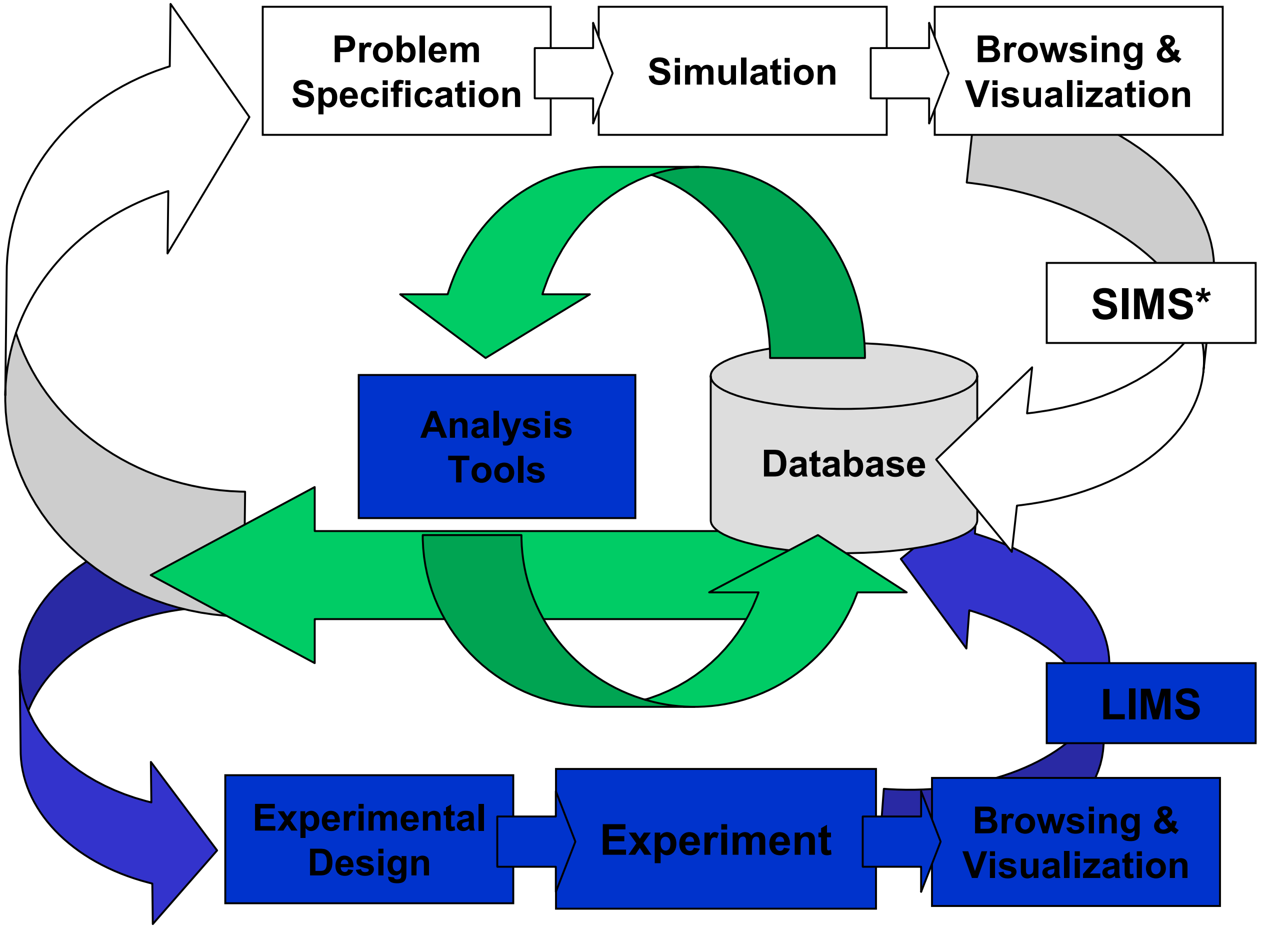
Adducted DNA

DNA-enzyme complex

ACTGGCATTCA—TATAGCT

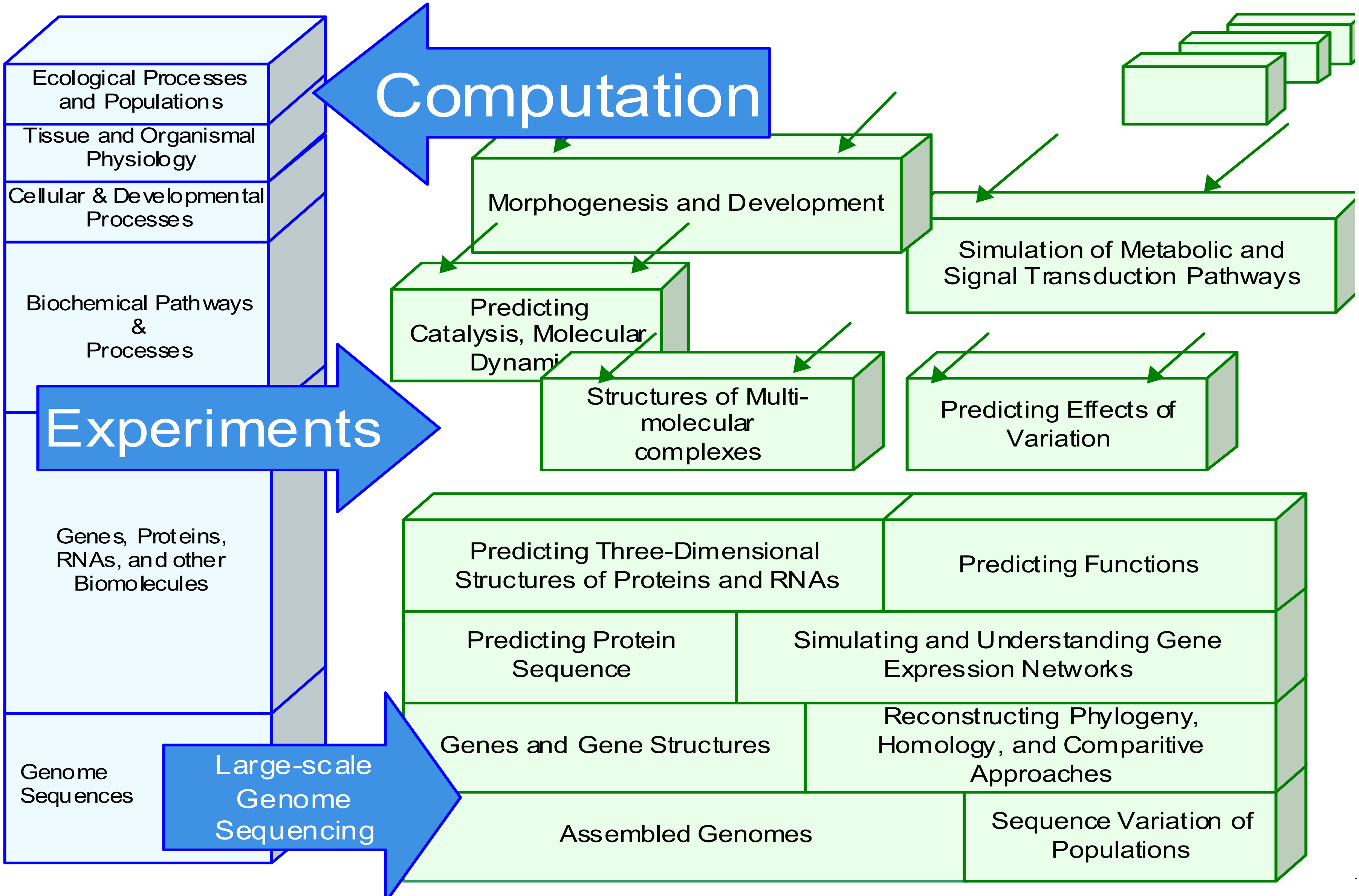
From Mike Colvin, LLNL

An Integrated View of Simulation, Experiment, and Bioinformatics



*Simulation Information Management System

Genomics is Powering the New Biology, but Computing is in the Drivers Seat



Ecological Processes and Populations

Tissue and Organismal Physiology

Cellular & Developmental Processes

Biochemical Pathways & Processes

Genes, Proteins, RNAs, and other Biomolecules

Genome Sequences

Computation

Experiments

Large-scale Genome Sequencing

Morphogenesis and Development

Simulation of Metabolic and Signal Transduction Pathways

Predicting Catalysis, Molecular Dynami

Structures of Multi-molecular complexes

Predicting Effects of Variation

Predicting Three-Dimensional Structures of Proteins and RNAs

Predicting Functions

Predicting Protein Sequence

Simulating and Understanding Gene Expression Networks

Genes and Gene Structures

Reconstructing Phylogeny, Homology, and Comparitive Approaches

Assembled Genomes

Sequence Variation of Populations

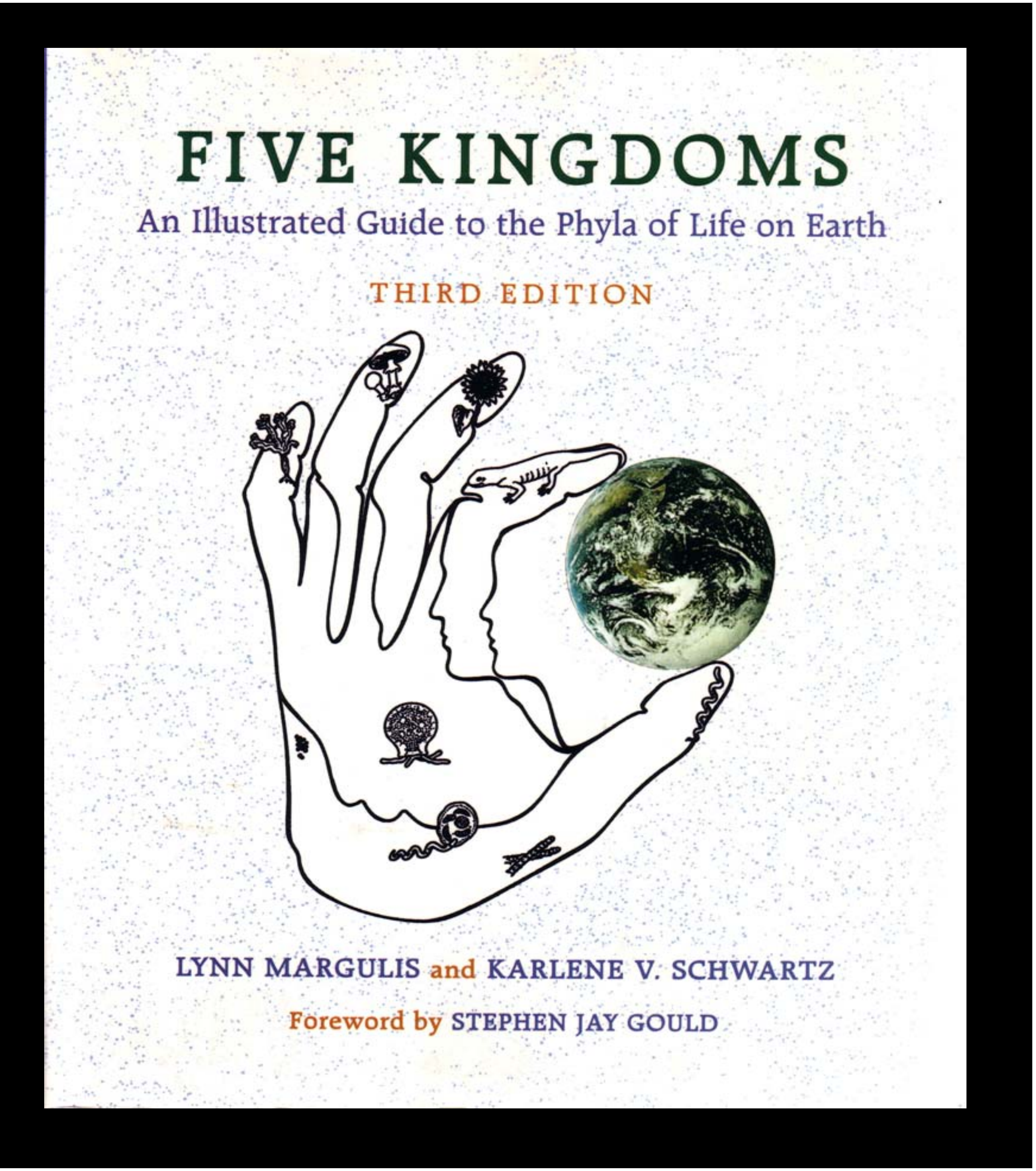
cago

Systems Biology

- Integrative (synthetic) understanding of a biological system
  - Cell, organism, community and ecosystem
- Counterpoint to reductionism
  - Requires synthesizing knowledge from multiple levels of the system
- Discovery oriented not necessarily hypothesis driven
  - Data mining vs theorem proving

Biology is BIG!!

- 3500 Millions years of evolution
  - ~10M extant species (distinct genomes)
    - 200 genomes sequenced so far many to go!!
    - >100M extinct species
  - 10^8 average genome size (coding region)
  - ~10^8 bp x ~10^7 sp = 10^{15} bp of genetic diversity
  - 10^{11}-10^{13} total genes and gene products
    - ~2,000 protein structures determined
    - typical bacteria has about 3,000 types of proteins



FIVE KINGDOMS

An Illustrated Guide to the Phyla of Life on Earth

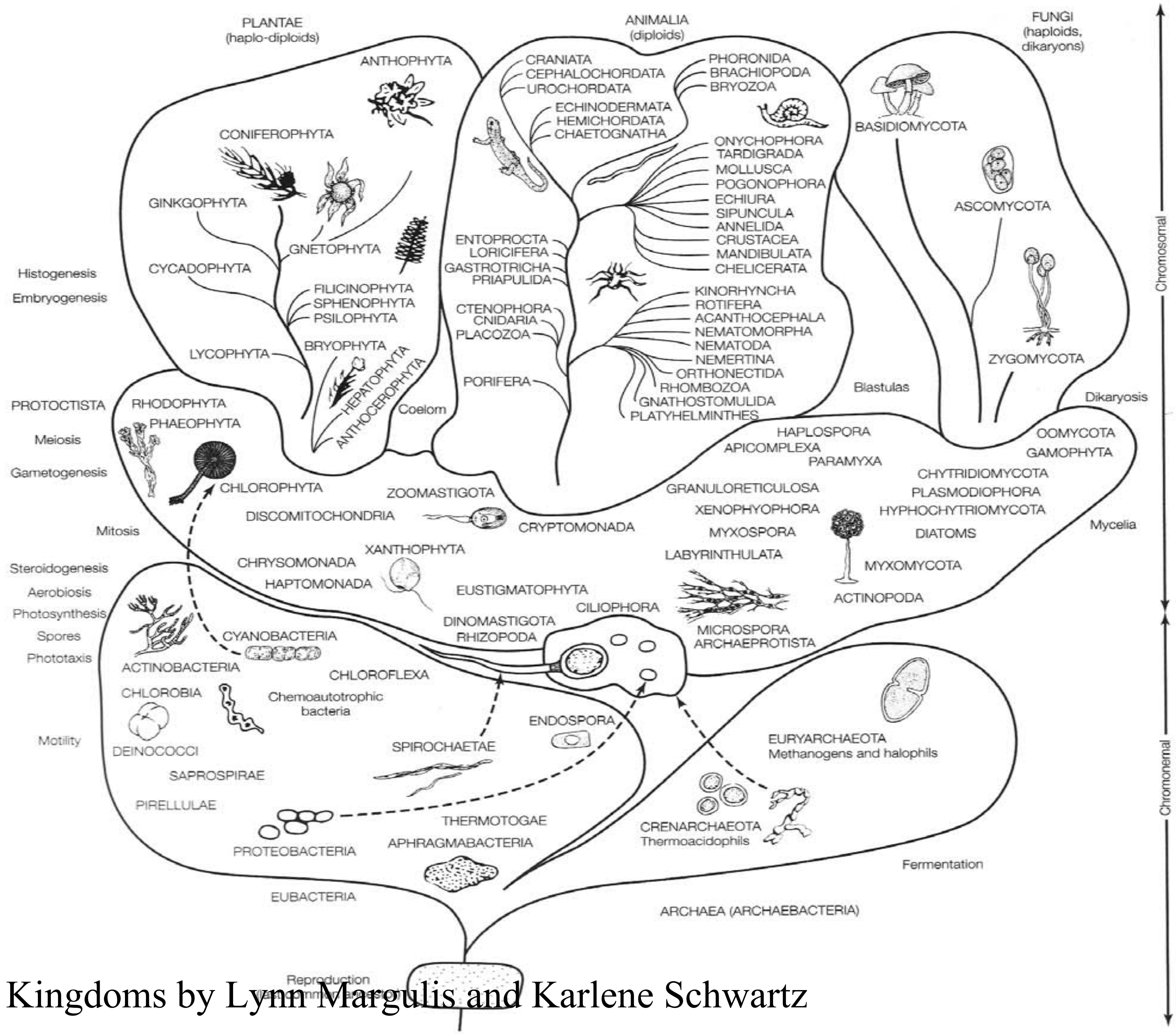
THIRD EDITION

LYNN MARGULIS and KARLENE V. SCHWARTZ

Foreword by STEPHEN JAY GOULD

- Overview of Life on Earth
- Basic Microbiology Introduction
- Excellent Summary

The phyla of life on Earth based on our modification of the Whittaker five-kingdom system and the symbiotic theory of the origin of eukaryotic cells.

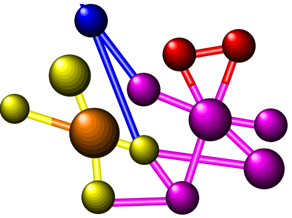


From Five Kingdoms by Lynn Margulis and Karlene Schwartz

From John Wooley

Biomedical Data:
High Complexity and Large Scale

billions



millions

Protein-Protein Interactions

metabolism
pathways
receptor-ligand
4° structure

Hundred thousands

Proteins

sequence
2° structure
3° structure

MPMILGYWDIRGLAHAIRLLLEYTDSSYEEKKYT...

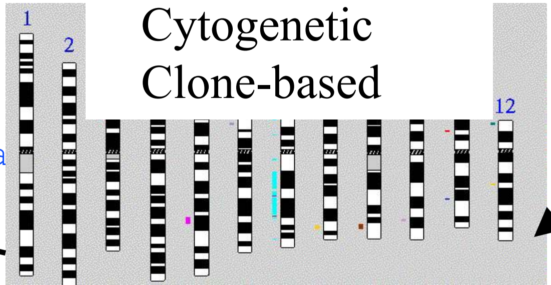
billions

DNA sequences

alignments

...atcgaattccaggcgtcacattctcaattcca

millions

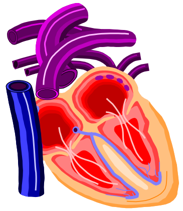


Genetics and Maps

Linkage
Cytogenetic
Clone-based

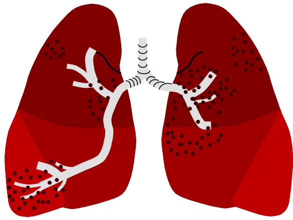
Polymorphism and Variants

genetic variants
individual patients
epidemiology

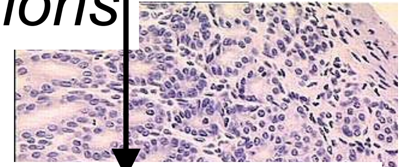


Physiology

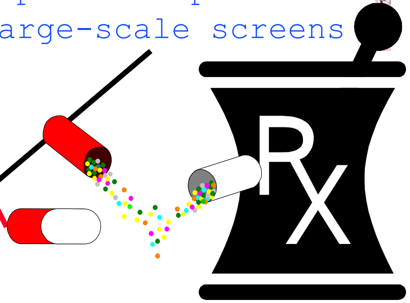
Cellular biology
Biochemistry
Neurobiology
Endocrinology
etc.



millions



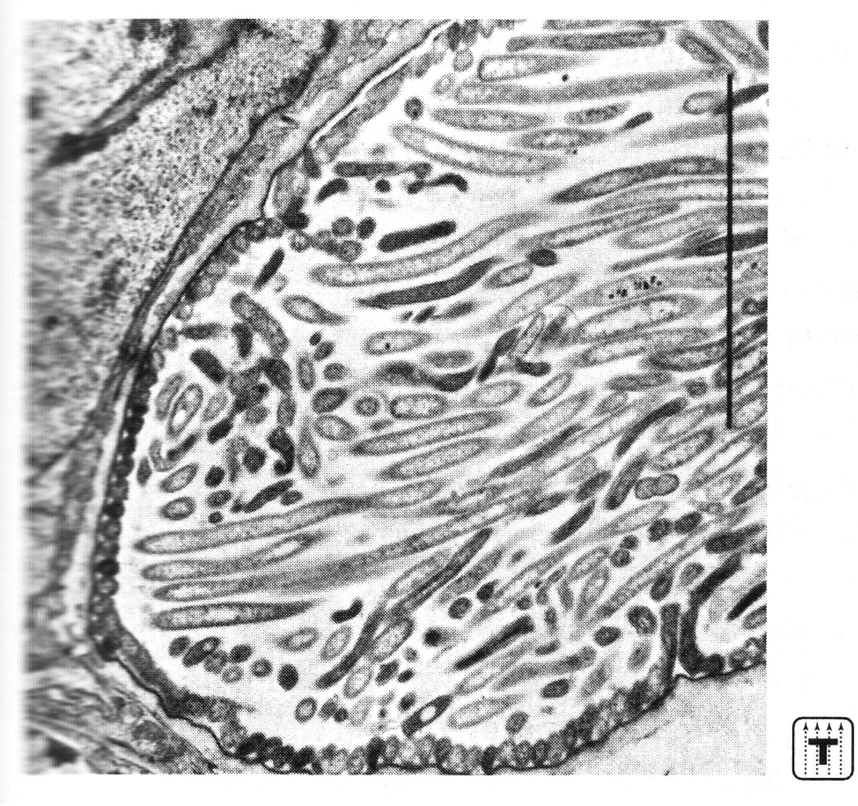
ESTs
Expression patterns
Large-scale screens



Will Biology Dominate the Grid?

- The largest science discipline:
  - The most scientists (Globally ~500,000-1,000,000)
  - The most research funding (Globally ~$50 Billion/year)
  - The most graduate students (>20,000 year)
- Strong couplings to:
  - Medicine and human health
  - Agriculture and food supplies
  - Energy and ecology
  - Future industrial processes (bio-nano)
  - Consumer of other scientific technologies

A Diverse Bacterial Community

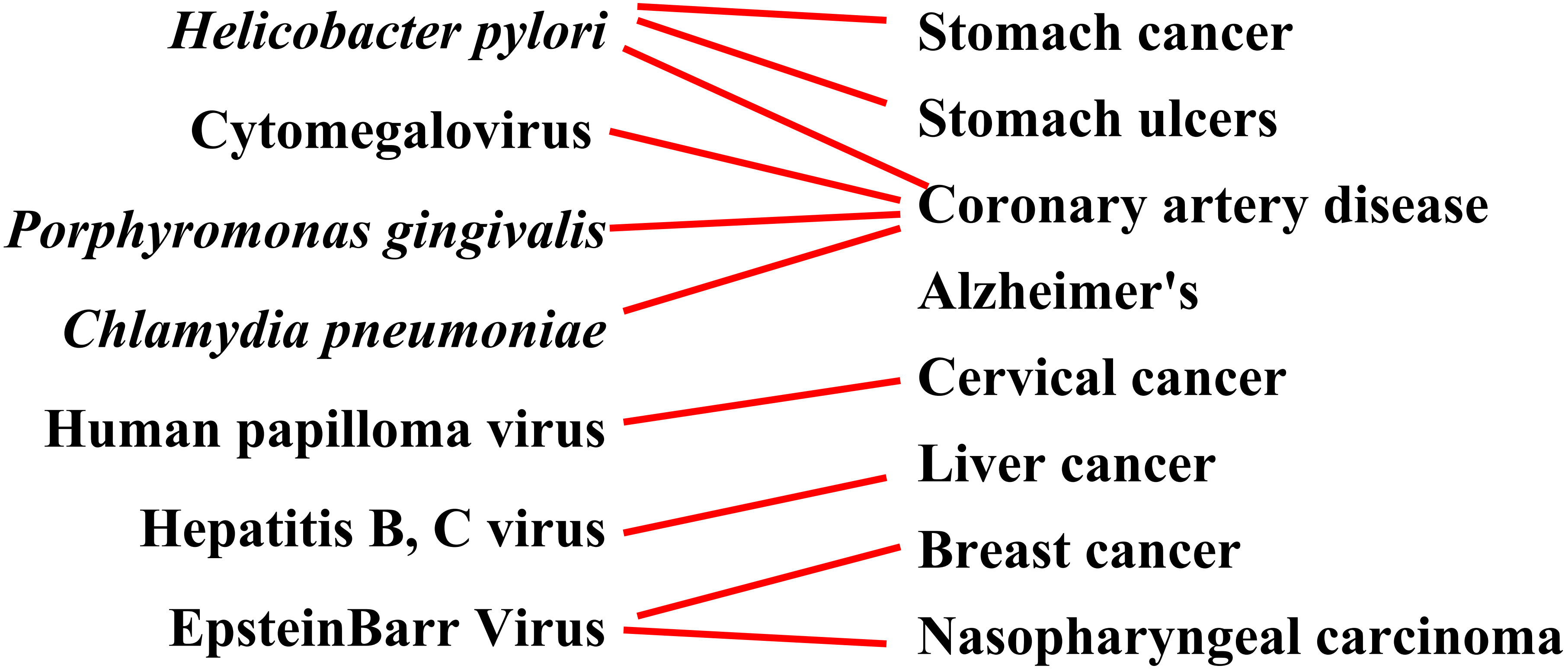


From Five Kingdoms

Lynn Margulis and Karlene Schwartz

- Pocket in the hindgut wall of the Sonoran desert termite *Pterotermes occidentis*
- 10 billion bacteria per milliliter
- Anoxic environment
- ~30 strains are facultative aerobes
- Many/most are unknown

Human Microbial Ecology: Sorting Out Cause, Effect + Cure



Helicobacter pylori

Cytomegalovirus

Porphyromonas gingivalis

Chlamydia pneumoniae

Human papilloma virus

Hepatitis B, C virus

EpsteinBarr Virus

Stomach cancer

Stomach ulcers

Coronary artery disease

Alzheimer's

Cervical cancer

Liver cancer

Breast cancer

Nasopharyngeal carcinoma

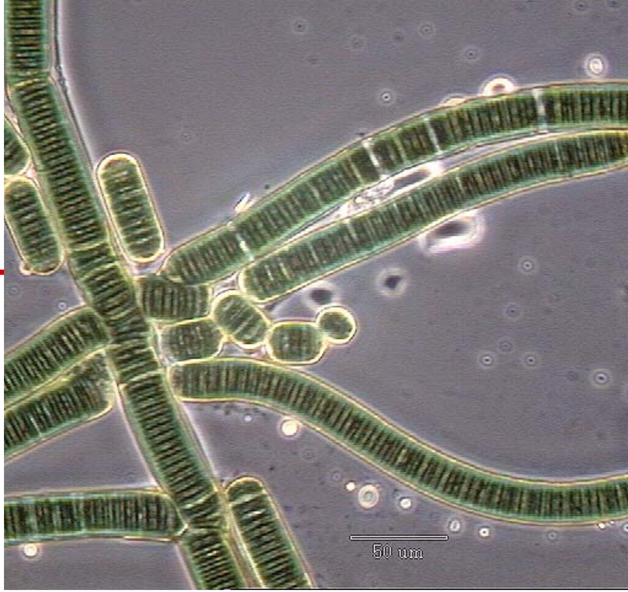
— = suspected linkages

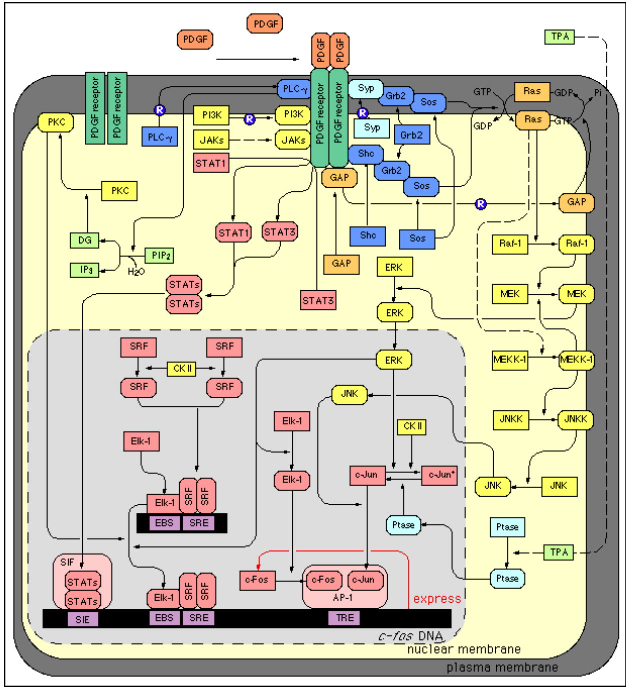
An Initial Focus on Prokaryotic life

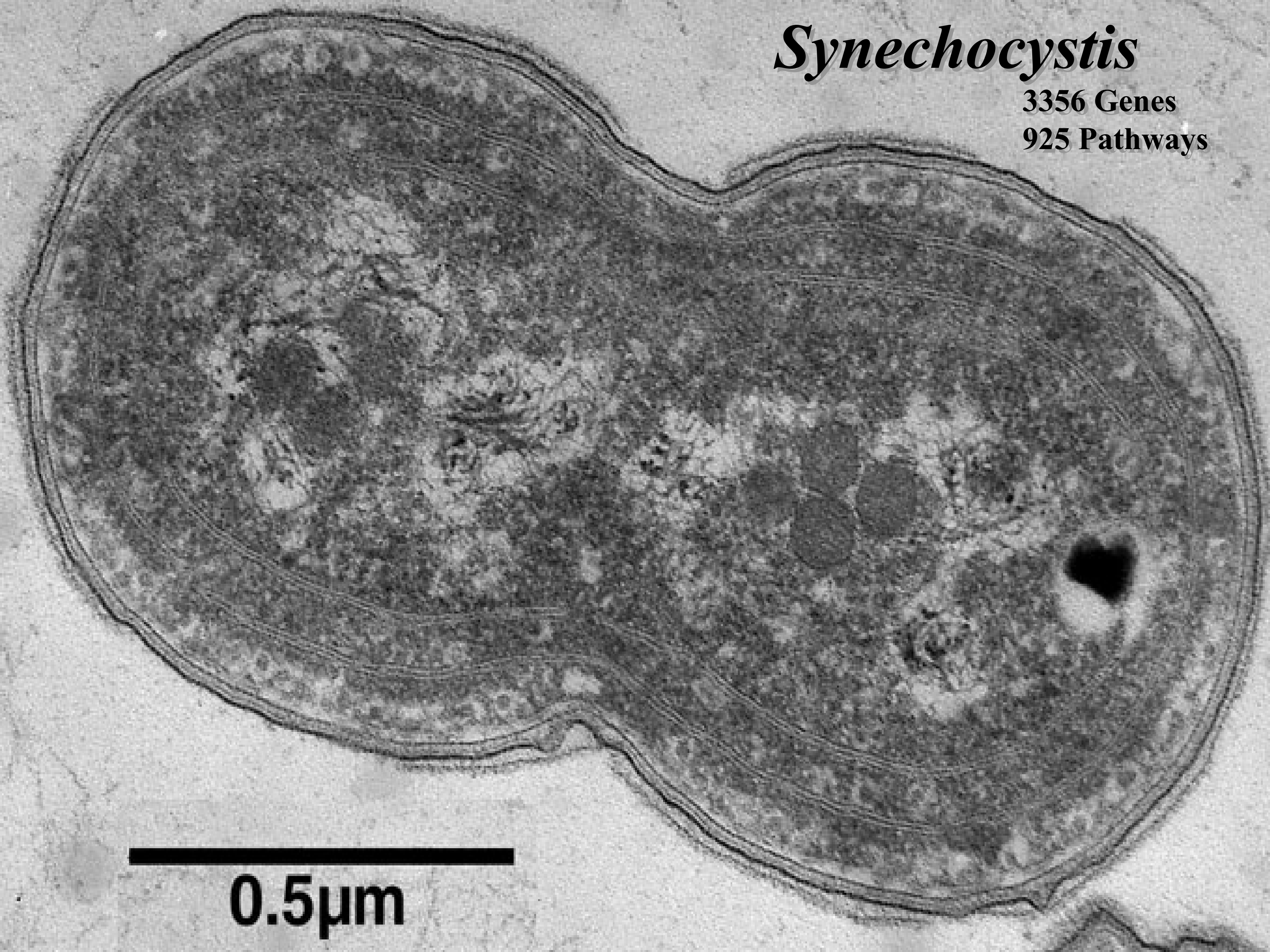
- ~5,000 known species of prokaryotic life
  - It is estimated that we have identified only 1%-5% of extant species ⇒ 80,000-400,000 species
- More diversity than Eukaryotic life forms
  - More diverse metabolisms, more diverse environments
- A Human contains 10^{12} cells and 10^{13} bacterial cells
  - We have little current understanding of human micro ecology

Biological CAD: Enabling BioDesign

- Understanding and manipulating biological systems from an information systems standpoint [e.g. organization, communication, transformation]
  - Genotype + Environment = Phenotype
  - Strong analogs to VLSI design tools
- Ultimately goal is designing new biological structures and systems
  - Custom microbe design
  - Metabolic Engineering
  - Synthetic model organisms







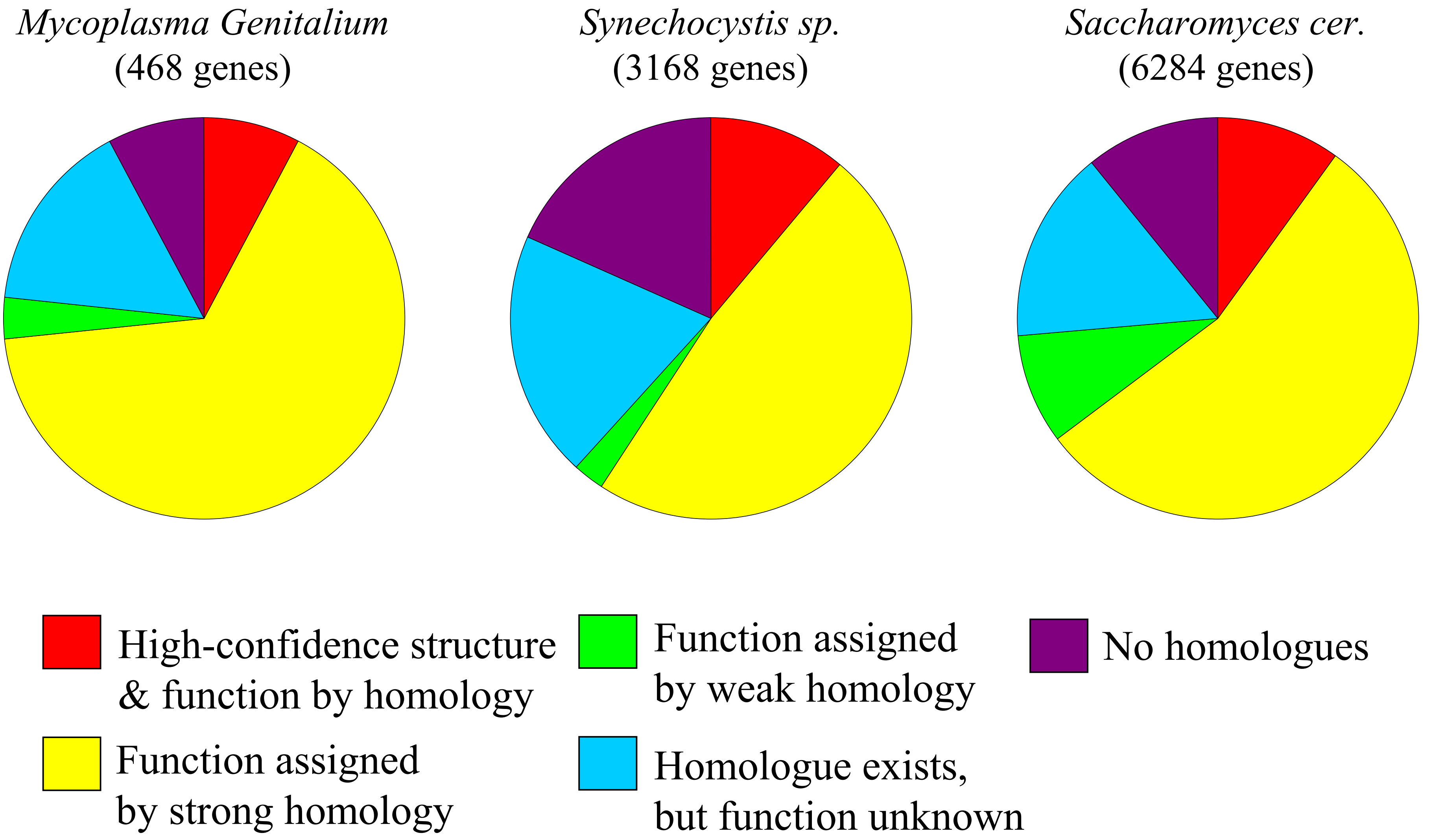
Synechocystis

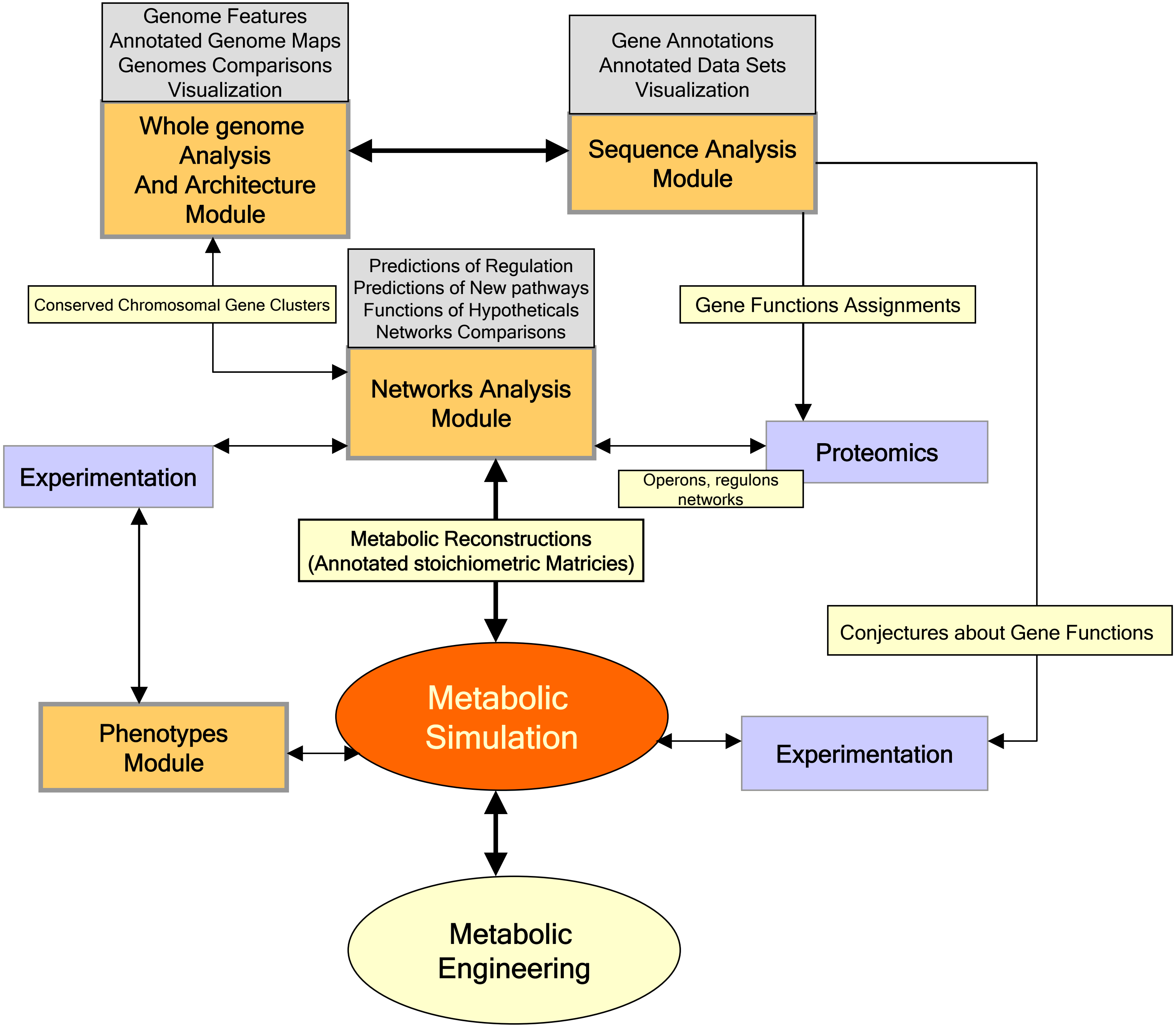
3356 Genes

925 Pathways

Degree of whole-genome structure & function assignment varies by organism (and required confidence level)

Results for three organisms using Genequiz (EBI)





Genome Features
Annotated Genome Maps
Genomes Comparisons
Visualization

Whole genome Analysis And Architecture Module

Gene Annotations
Annotated Data Sets
Visualization

Sequence Analysis Module

Conserved Chromosomal Gene Clusters

Predictions of Regulation
Predictions of New pathways
Functions of Hypotheticals
Networks Comparisons

Networks Analysis Module

Gene Functions Assignments

Experimentation

Proteomics

Operons, regulons networks

Metabolic Reconstructions (Annotated stoichiometric Matricies)

Conjectures about Gene Functions

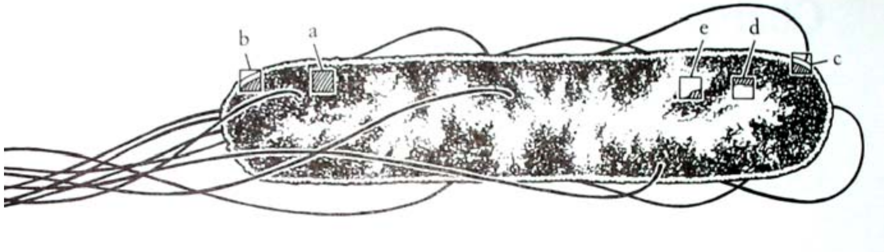
Phenotypes Module

Metabolic Simulation

Experimentation

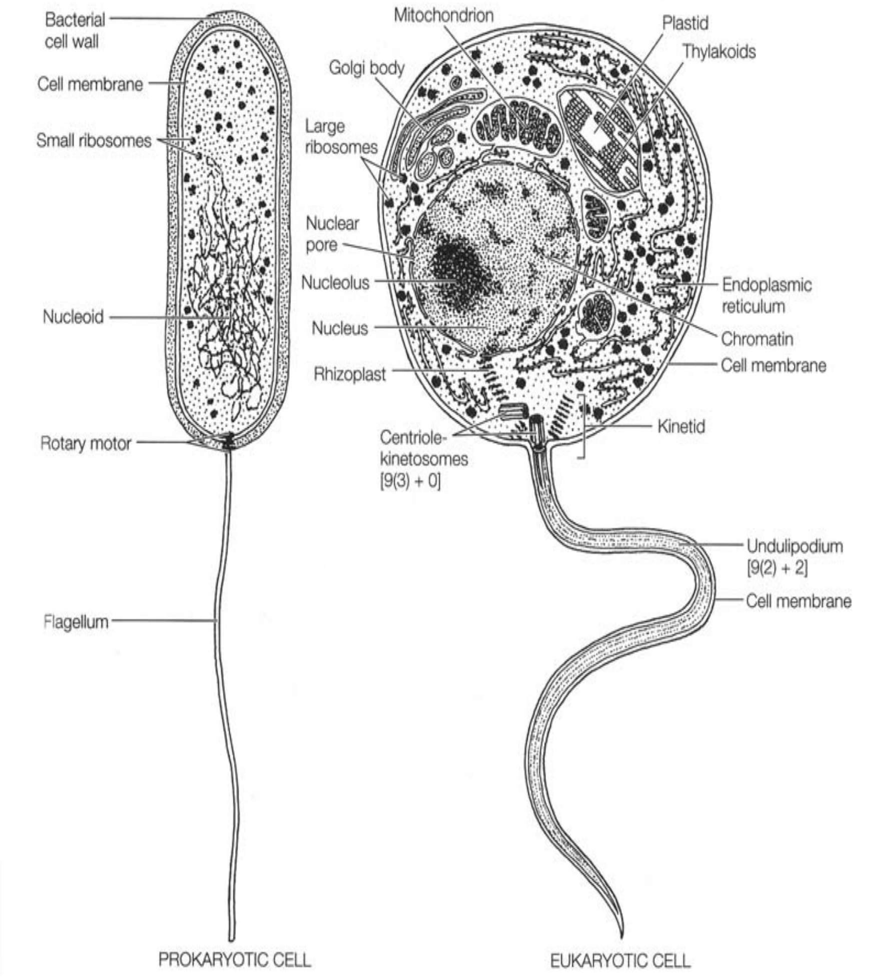
Metabolic Engineering

Reconstructing and Modeling a Prokaryotic Cell?



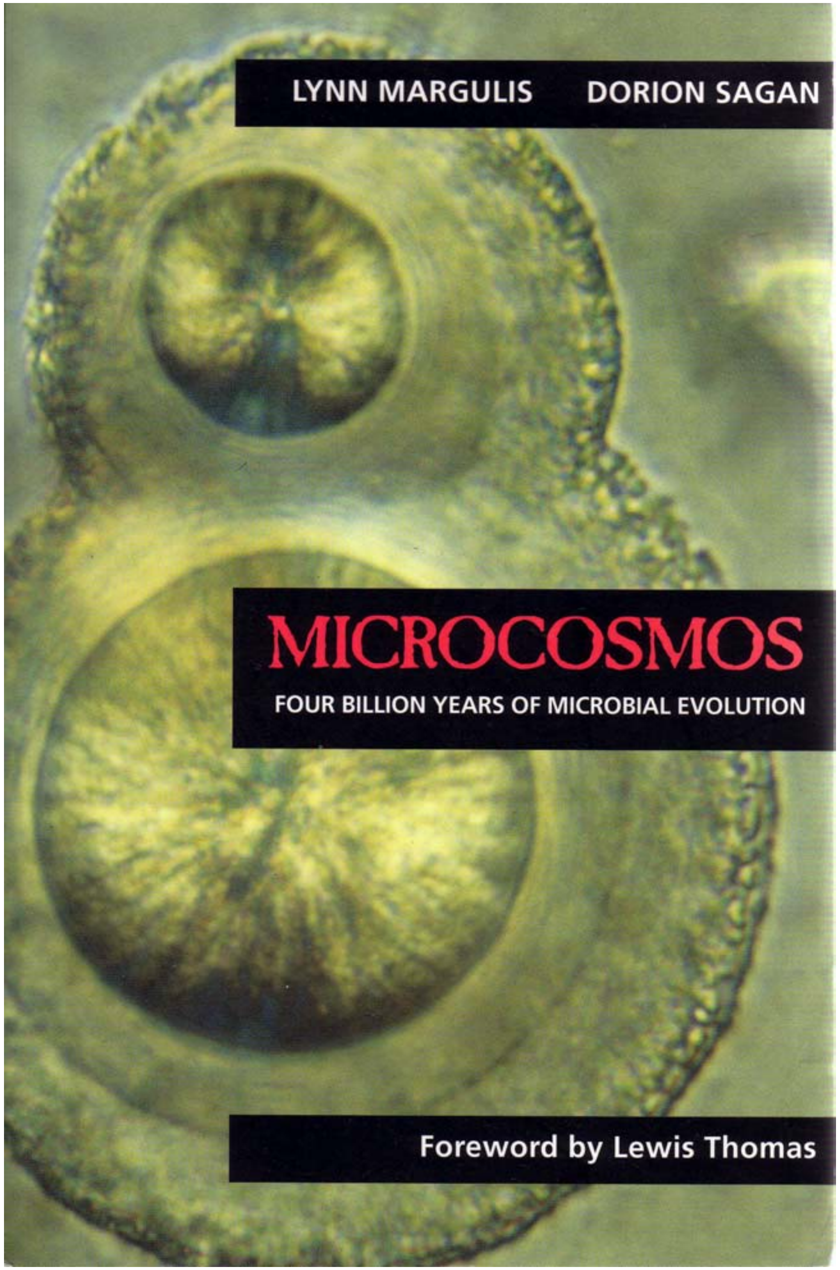
- Typical bacteria [*E. coli*]
  - 1000nm x 300nm x 300nm volume
  - ~4000 genes and gene products
    - 1/4 genes ⇒ protein synthesis
    - 1/4 genes ⇒ glycolysis
    - 1/4 genes ⇒ citric acid cycle
    - rest genes involved in regulation, synthesis and degrading tasks
    - Relatively few genes related to sensing and motility
  - 1,000's of small molecular species, not tracked individually
  - ~3 million total large molecules to track

Prokaryotic v Eukaryotic Cell



- Membrane separated nucleus
- Multiple cellular compartments
- Introns
- More complex cell membrane
- Complex motility mechanisms
- Etc.

From Lynn Margulis and Karlene Schwartz



LYNN MARGULIS DORION SAGAN

MICROCOSMOS

FOUR BILLION YEARS OF MICROBIAL EVOLUTION

Foreword by Lewis Thomas

- Lynn Margulis and Dorion Sagan
- Story of microbial evolution
- Introduction to early life on earth

Intracellular Environment – Gel Like Media

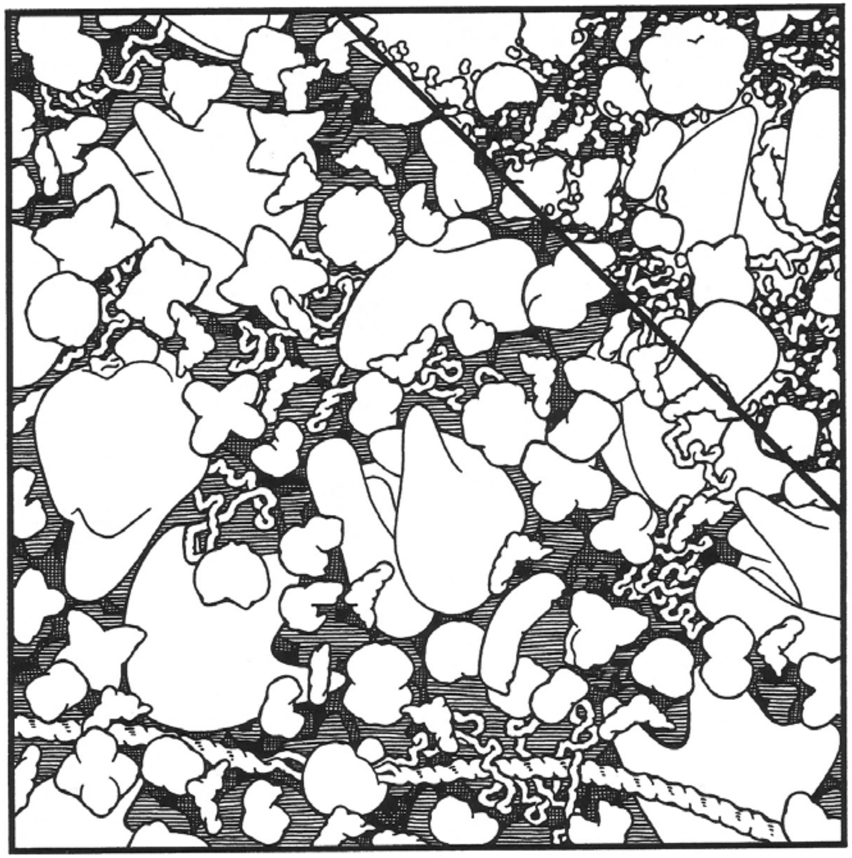


Figure 4.2 Cytoplasm

From: David Goodsell, The Machinery of Life

- 100 nm^3
- 450 proteins
- 30 ribosomes
- 340 tRNA molecules
- Several long mRNAs
- 30,000 small organic molecules
- 50,000 Ions
- Rest filled with water 70%

Cell Membranes and Cell Wall

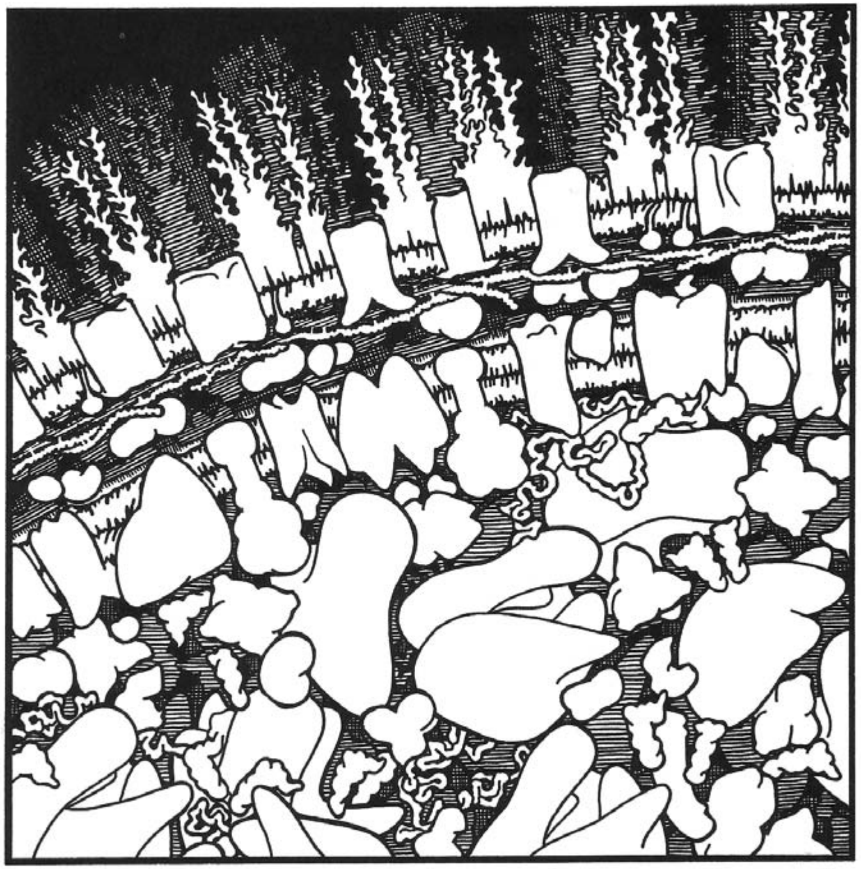


Figure 4.3 Cell Wall

From: David Goodsell, The Machinery of Life

- Cell wall
  - Polysaccharides
  - Porin pores
- Peptidoglycan
  - cross linked
- Periplasmic space
  - Small proteins
- Complex inner membrane
  - < 50% lipids

Flagellum and Flagellar Motor: Nanotechnology

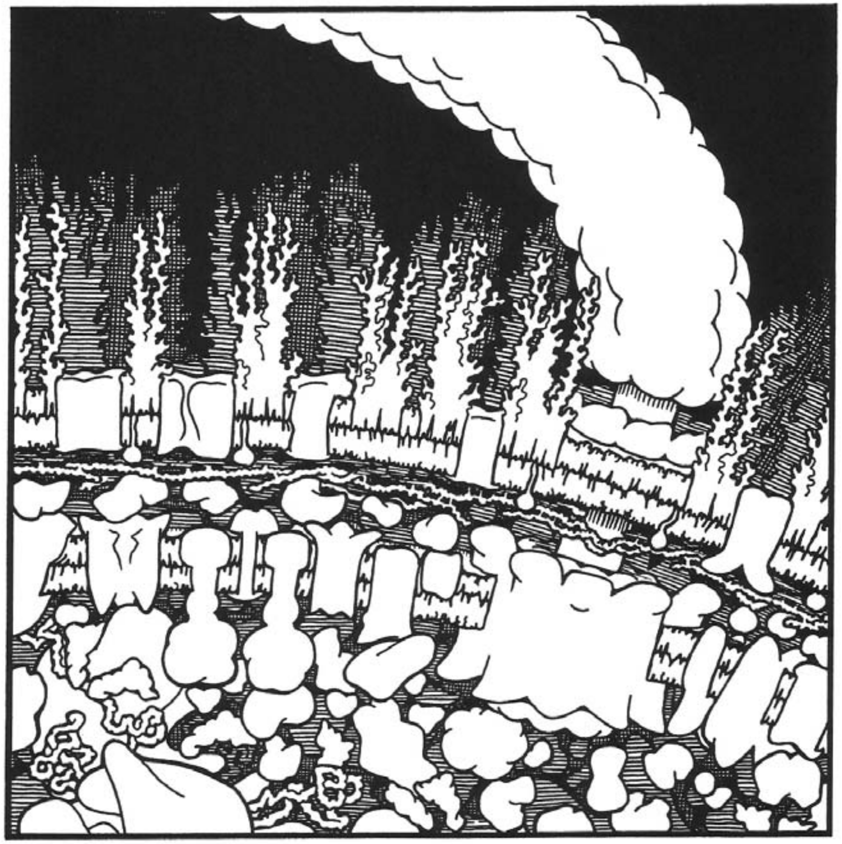


Figure 4.4 Flagellum and Flagellar Motor

From: David Goodsell, The Machinery of Life

- Transmembrane proton powered rotating motor
- About 10 Flagella per cell
- 5-10 um long
  - Built from the inside out
- Propels cell ~10-20 um/sec
  - Medium is extremely viscous
  - 10-20 body lengths/sec
  - ~100KM/hr scaled velocity

DNA Replication via DNA Polymerase

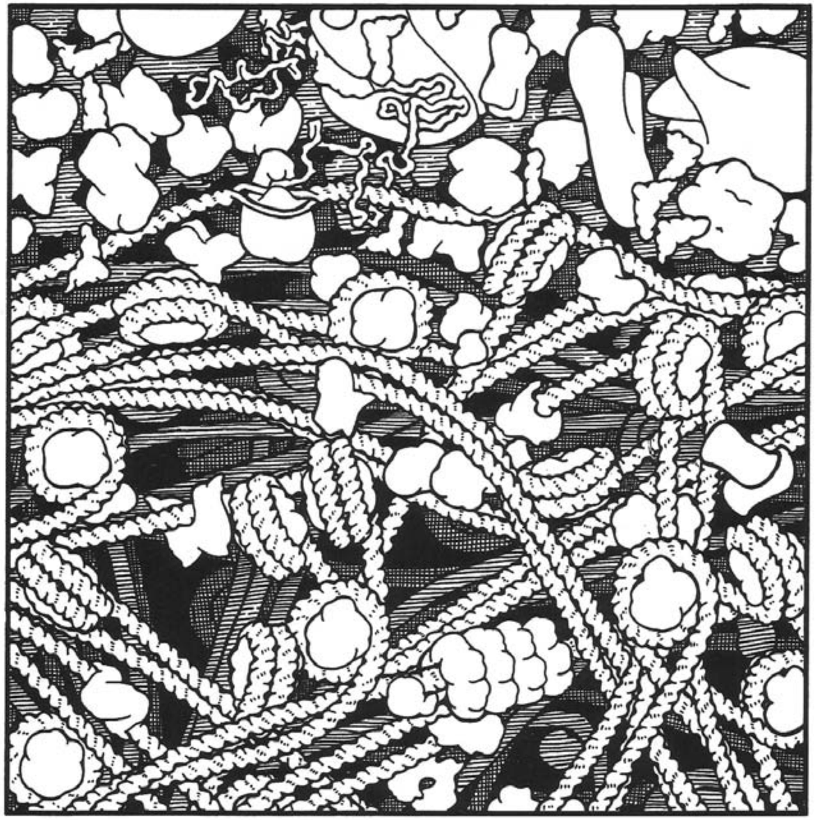
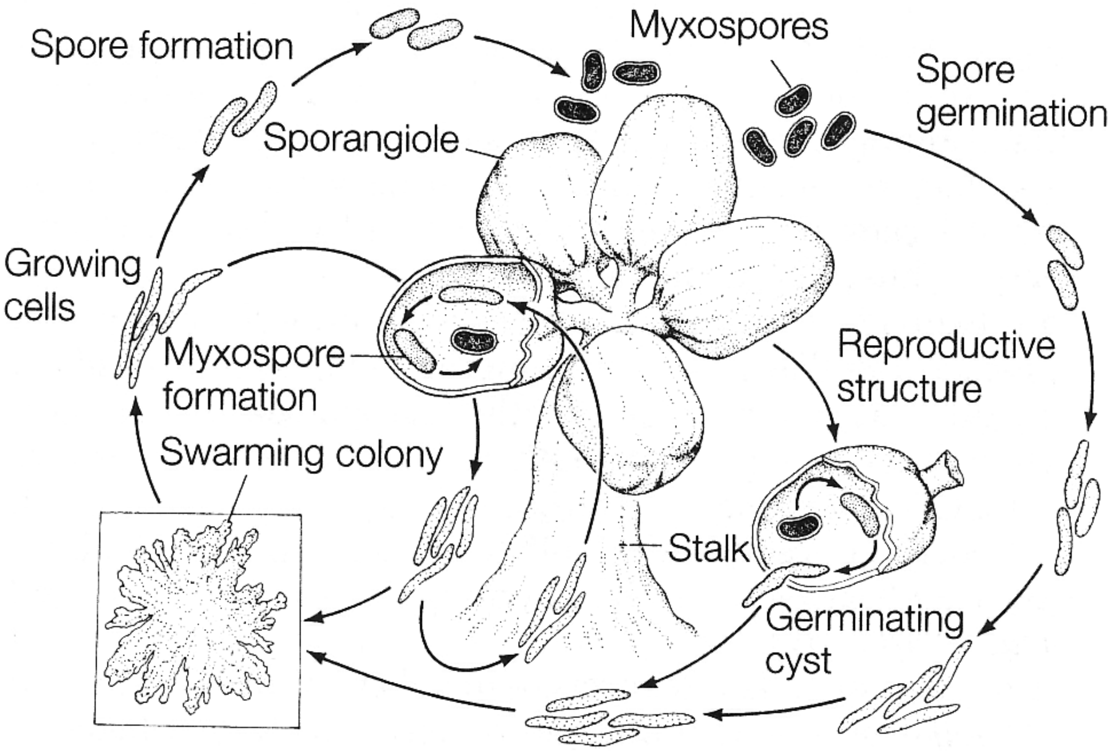


Figure 4.5 Nuclear Region

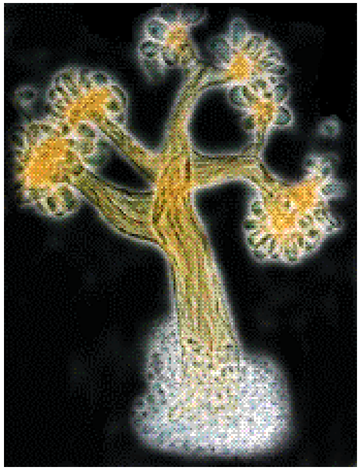
- DNA replication about 800 new nucleotides per second
- In circular DNA both directions at once
- 50 minute to duplicate entire circle of 4,700,000 nucleotides
- With cell replication ~30 minutes
- DNA replication is pipelined!!

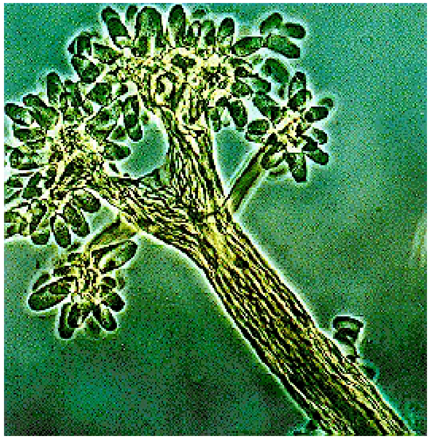
From: David Goodsell, The Machinery of Life

Understanding Bacterial Life Cycles



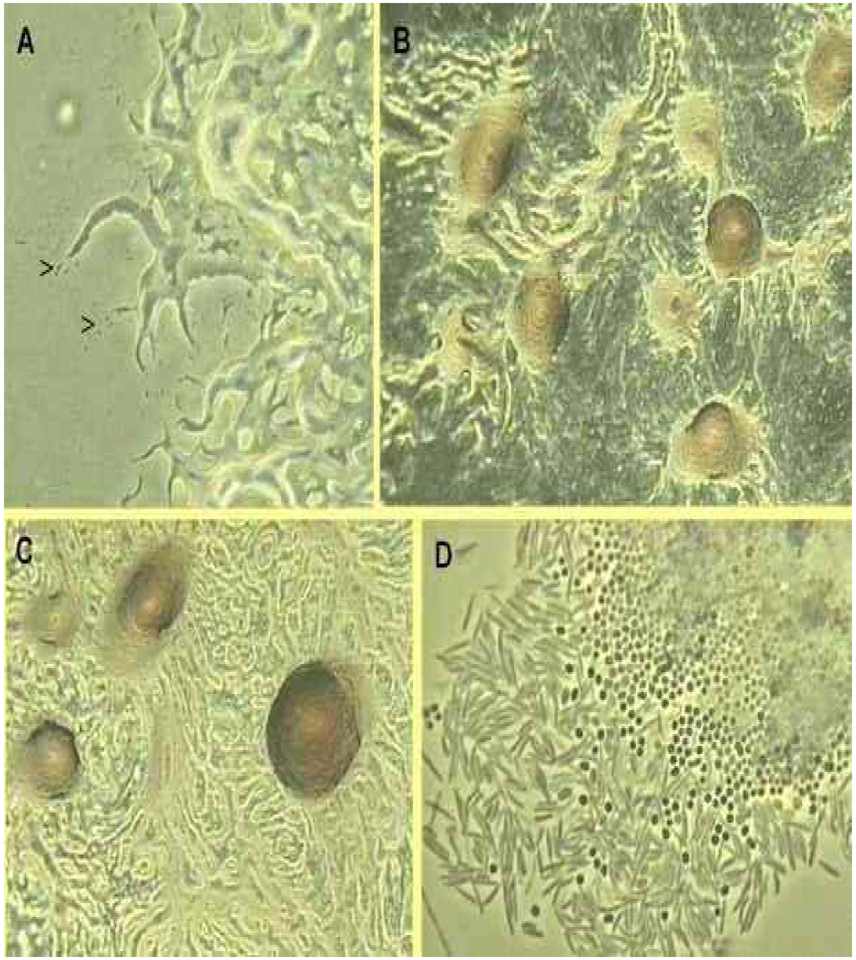
F Life cycle of *Stigmatella aurantiaca.* [Drawing by L. Meszoly; labeled by M. Dworkin.]





From Lynn Margulis and Karlene Schwartz

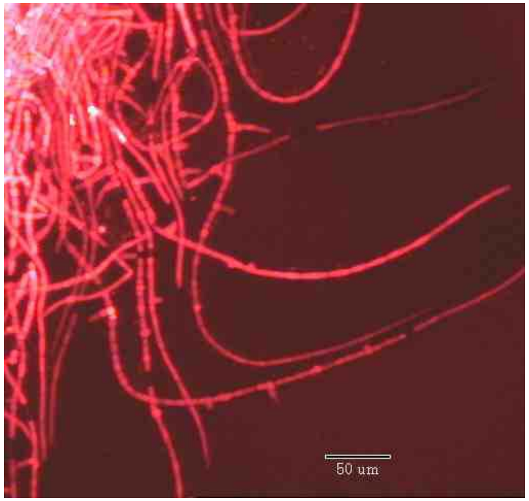
Modeling Swarming Behavior in *Myxobacteria*



- 100,000 cells swarm to form fruiting bodies
- 80% of the cells lyse
- 20% form spores
- Involves chemotaxis and quorum sensing
- Most complex bacterial genome currently known at > 9Mbp
- Very little is understood

MONERA: Hierarchical Biological System Modeling Environment

- Genetic Sequences
- Molecular Machines
- Molecular Complexes and modules
- Networks + Pathways [metabolic, signaling, regulation]
- Structural components [ultrastructures]
- Cell Structure and Morphology
- Extracellular Environment
- Populations and Consortia etc.



Systems Biology Model Development and Sharing

Simulators	Director	Institution	Features
ERATO,j	John Doyle	Caltech	planned workbench
Gepasi,w	Pedro Mendes	Santa Fe	MCA, systems kinetics
JarnacScamp,wx	Herbert Sauro	Caltech	MCA, Stochastic
StochSim,w+	Dennis Bray	Cambridge	Stochastic
BioSpice,u	Adam Arkin	LBL	Stochastic
DBSolve,w	Igor Goryanin	Glaxo	enzyme/receptor-ligand
E-Cell,u+	Masaru Tomita	Keio	metabolism. Net ODE
Vcell,j	Jim Schaff	U.CT	geometry
Xsim,u	J.Bassingthwaighte	Seattle	enzymes to body physiology
CellML,x+	Peter Hunter	U.Auckland	geometry, model sharing
GENESIS,u	James Bower	Caltech	neural networks
Simex,u+	Lael Gatewood	U.MN	Stochastic micro populations
MONERA, ux+	Selkov/Stevens	ANL/UC	multilevel ODE/Logic models

j=java, w= windows, u=unix, x=XML, + = source/community input

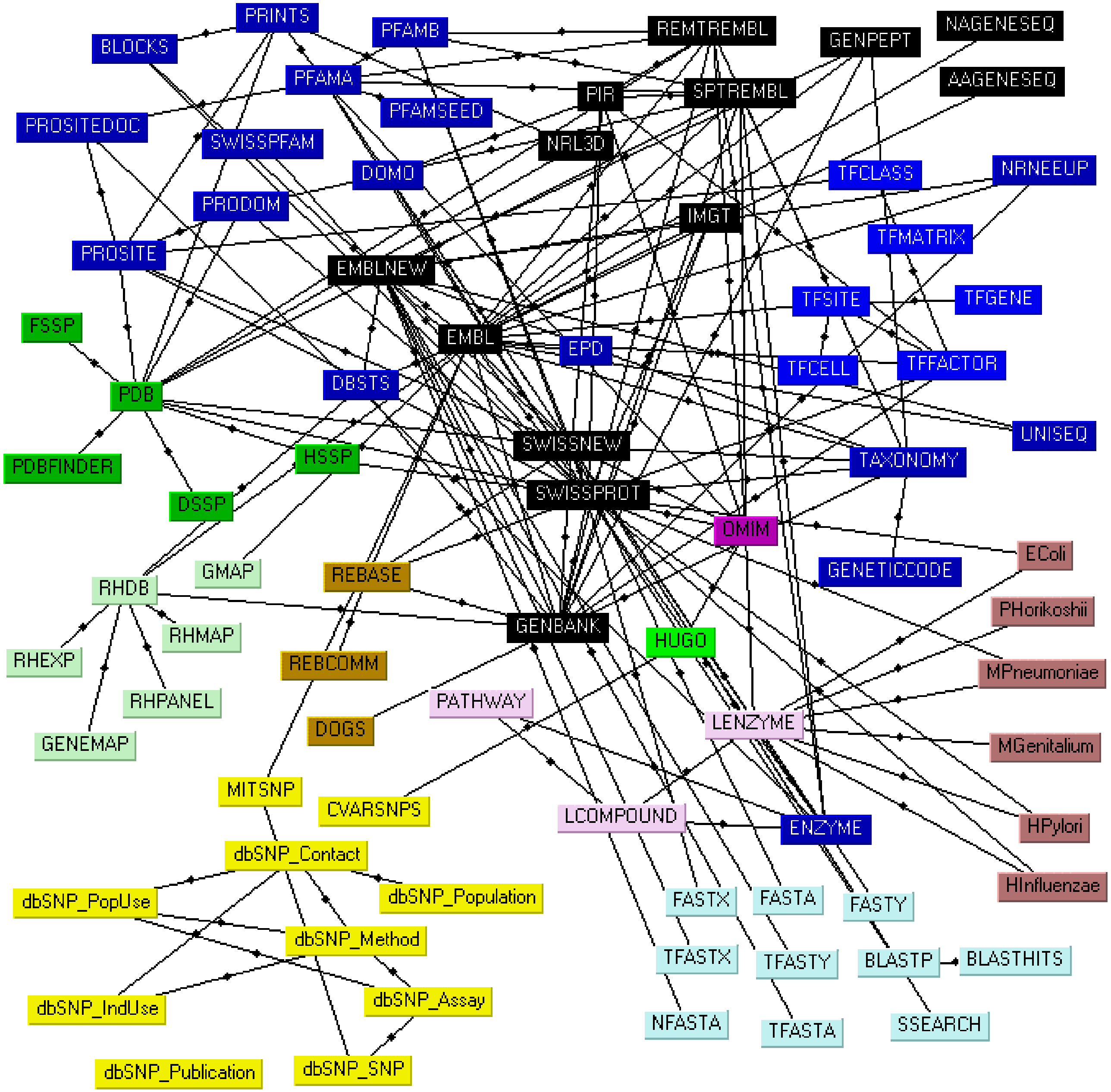
What the BioGrid Needs To Provide?

- Scalable compute and data capabilities beyond that available locally
- Distributed infrastructure available 24x7 worldwide
- Integration with local bioinfo systems for seamless computing and data management
- Enables leverage of remote systems administration and support via service providers
- Enables access to state of the art facilities at fraction of the cost (SPs just add more servers)
- Centralized support of tools and data
- Bottom line ⇒ <u>enables biologists to focus on biology</u>

Biology Databases (335 in 2001)

- Major Seq. Repositories (7)
- Comparative Genomics (7)
- Gene Expression (19)
- Gene ID & Structure (31)
- Genetic & Physical Maps (9)
- Genomic (49)
- Intermolecular Interactions (5)
- Metabolic Pathways & Cellular Regulation (12)
- Mutation (34)
- Pathology (8)
- Protein (51)
- Protein Sequence Motifs (18)
- Proteome Resources (8)
- Retrieval Systems & DB Structure (3)
- RNA Sequences (26)
- Structure (32)
- Transgenics (2)
- Varied Biomedical (18)

Baxevanis, A.D. 2002. *Nucleic Acids Research* 30: 1-12.

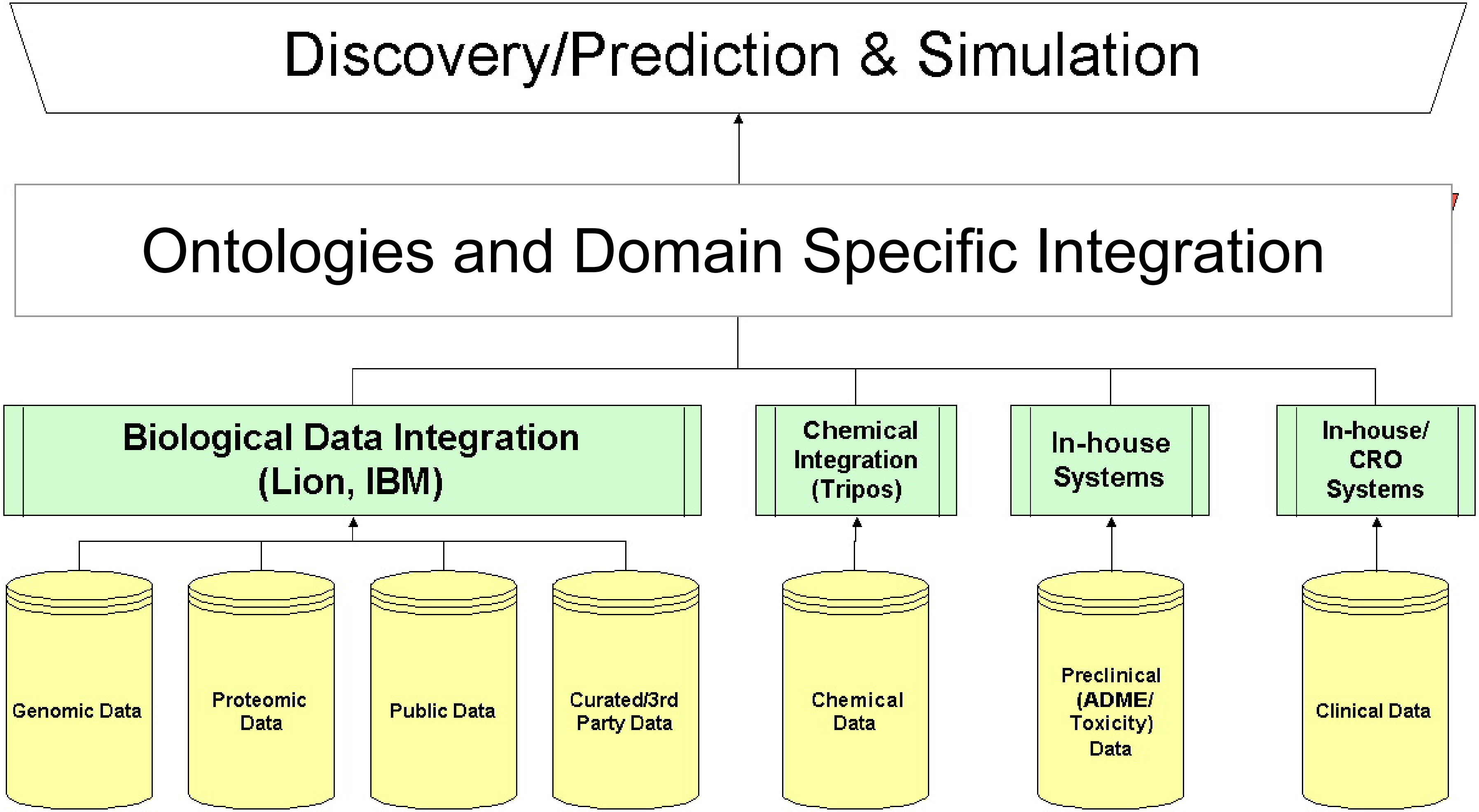

PRINTS
BLOCKS
PFAMB
PFAMA
PFAMSEED
REMTREMBL
GENPEPT
NAGENESEQ
AAGENESEQ
PIR
SPTREMBL
PROSITEDOC
SWISSPFAM
NRL3D
DOMO
PRODOM
TFCLASS
NRNEEUP
IMGT
PROSITE
EMBLNEW
TFMATRIX
TFSITE
TFGENE
FSSP
EMBL
EPD
TFCELL
TFFACTOR
PDB
DBSTS
UNISEQ
SWISSNEW
TAXONOMY
PDBFINDER
HSSP
SWISSPROT
DSSP
OMIM
ECoIi
GENETICCODE
GMAP
RHDB
REBASE
PHorikoshii
GENBANK
HUGO
RHMAP
RHEXP
REBCOMM
MPneumoniae
PATHWAY
RHPANEL
LENZYME
DOGS
MGenitalium
GENEMAP
MITSNP
CVARSNPS
LCOMPOUND
ENZYME
HPylori
dbSNP_Contact
dbSNP_PopUse
dbSNP_Population
HInfluenzae
FASTX
FASTA
FASTY
dbSNP_Method
TFASTX
TFASTY
BLASTP
BLASTHITS
dbSNP_IndUse
dbSNP_Assay
NFASTA
TFASTA
SSEARCH
dbSNP_Publication
dbSNP_SNP

LION

Grids vs Web tools for biology

- The biology community has developed an extensive collection of web resources to support research:
  - Databases and search engines (entrez, etc)
  - Functional annotation systems (wit, etc.)
  - Organism specific databases (ecocyc, etc.)
  - Literature search engines (pubmed, etc.)
  - Web based modeling systems (vcell, etc.)

Software Infrastructure in Drug Discovery



Discovery/Prediction & Simulation

Ontologies and Domain Specific Integration

Biological Data Integration (Lion, IBM)

Chemical Integration (Tripos)

In-house Systems

In-house/ CRO Systems

Genomic Data

Proteomic Data

Public Data

Curated/3rd Party Data

Chemical Data

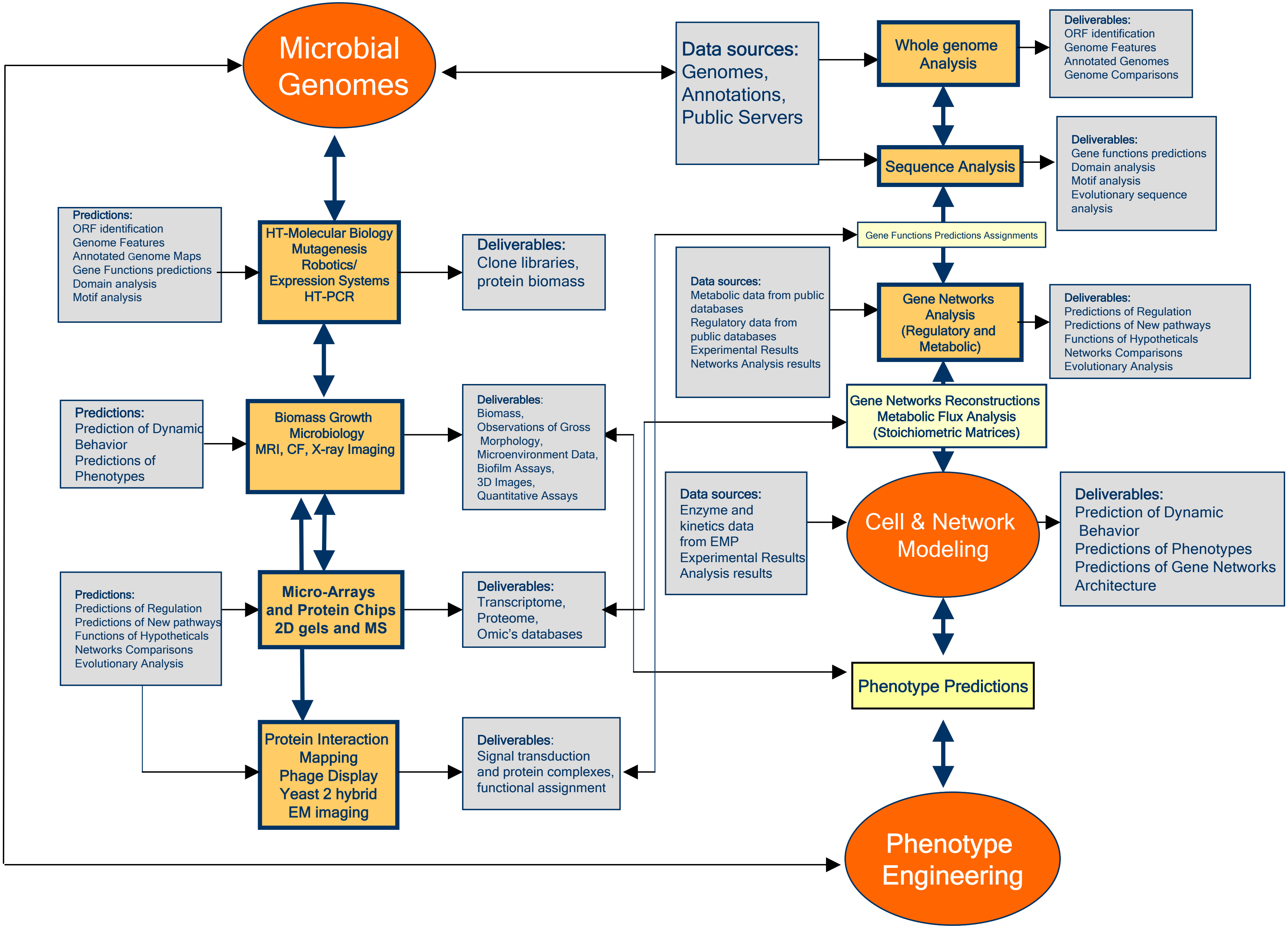
Preclinical (ADME/ Toxicity) Data

Clinical Data

From Richard Gardner(InCellico)

Requirements for the BioGrid

- Open and extendable architecture
  - Enable tie in to service stack at appropriate points
  - Not just access via Portals
- Leverage scripting tools in wide use for Bioinformatics
  - Create BioGrid services bindings for PERL and Python
- Address data federation and integration
  - Leverage work of IBM, Lion, etc.
- Match the biology workflow and tool chain
  - Create high-level BioGrid services to address critical stages in existing workflow
  - Support composibility of new BioGrid tools with existing tool chain elements



Microbial Genomes

Data sources:
Genomes,
Annotations,
Public Servers

Whole genome Analysis

Deliverables:
ORF identification
Genome Features
Annotated Genomes
Genome Comparisons

Sequence Analysis

Deliverables:
Gene functions predictions
Domain analysis
Motif analysis
Evolutionary sequence analysis

Gene Functions Predictions Assignments

Predictions:
ORF identification
Genome Features
Annotated Genome Maps
Gene Functions predictions
Domain analysis
Motif analysis

HT-Molecular Biology
Mutagenesis
Robotics/
Expression Systems
HT-PCR

Deliverables:
Clone libraries,
protein biomass

Data sources:
Metabolic data from public databases
Regulatory data from public databases
Experimental Results
Networks Analysis results

Gene Networks Analysis
(Regulatory and Metabolic)

Deliverables:
Predictions of Regulation
Predictions of New pathways
Functions of Hypotheticals
Networks Comparisons
Evolutionary Analysis

Gene Networks Reconstructions
Metabolic Flux Analysis
(Stoichiometric Matrices)

Predictions:
Prediction of Dynamic
Behavior
Predictions of
Phenotypes

Biomass Growth
Microbiology
MRI, CF, X-ray Imaging

Deliverables:
Biomass,
Observations of Gross Morphology,
Microenvironment Data,
Biofilm Assays,
3D Images,
Quantitative Assays

Data sources:
Enzyme and kinetics data from EMP
Experimental Results
Analysis results

Cell & Network Modeling

Deliverables:
Prediction of Dynamic Behavior
Predictions of Phenotypes
Predictions of Gene Networks Architecture

Predictions:
Predictions of Regulation
Predictions of New pathways
Functions of Hypotheticals
Networks Comparisons
Evolutionary Analysis

Micro-Arrays
and Protein Chips
2D gels and MS

Deliverables:
Transcriptome,
Proteome,
Omic's databases

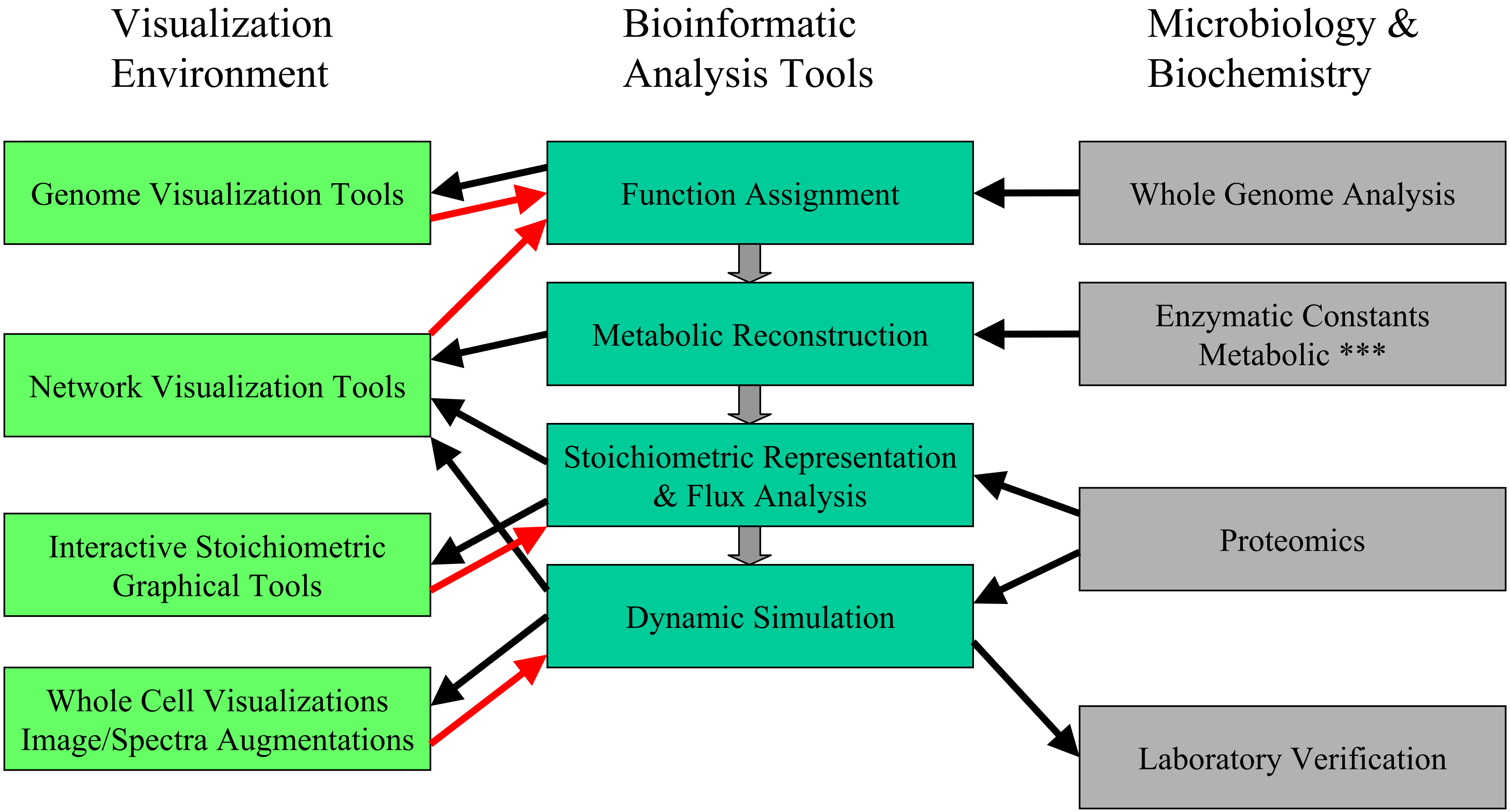
Phenotype Predictions

Protein Interaction Mapping
Phage Display
Yeast 2 hybrid
EM imaging

Deliverables:
Signal transduction and protein complexes, functional assignment

Phenotype Engineering

Visualization + Bioinformatics



Visualization Environment

Bioinformatic Analysis Tools

Microbiology & Biochemistry

Genome Visualization Tools

Network Visualization Tools

Interactive Stoichiometric Graphical Tools

Whole Cell Visualizations Image/Spectra Augmentations

Function Assignment

Metabolic Reconstruction

Stoichiometric Representation & Flux Analysis

Dynamic Simulation

Whole Genome Analysis

Enzymatic Constants Metabolic ***

Proteomics

Laboratory Verification

Anaplerotic reactions
PHOSPHOENOLPYRUVATE CARBOXYLASE
PYRUVATE CARBOXYLASE
OXALOACETATE
ACETYL-CoA
S(=L)-MALATE
SUCCINYL-CoA
GLYCINE
PROTEINS
File View Style Color Option Help
Glyoxylate cycle
OXALYL CoA
OXALATE
GLYOXYLATE
2-OXO-GLUTARATE
ISOCITRATE LYASE
OXALO-SUCCINATE
ACONITATE
ITACONATE
Citrate cycle
L-AMINO ACID
Align Options Help

Some BioGrid Challenges

- Scalable human bioinformatics expertise
  - Best people working on the important problems
  - Exploit collaboration technology to create world class teams
- Robust local bioinformatics computing environment
  - Best systems administrators and high-end technologies
  - Embed local resources into the Grid via portal technologies
- Access to leading edge bioinformatics software and databases customized to user needs
  - Core content from top scientists and developers
  - Integrated access to biological databases
- Worldwide access to robust computing and database infrastructure
  - Leverage Grid technology to provide worldwide access
  - Integrate purpose built systems and service providers

What We Need to Create

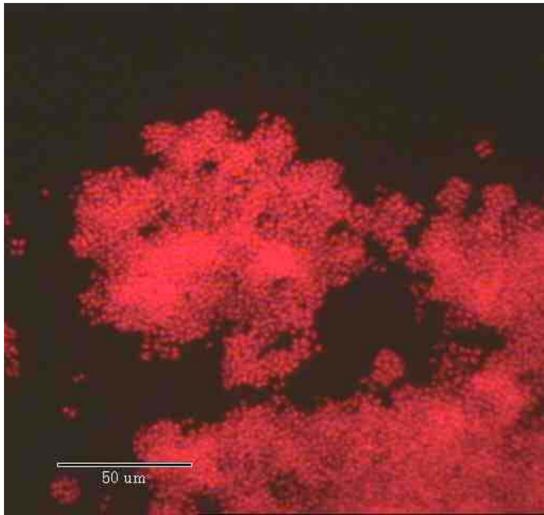
- Grid Bio applications enablement software layer
  - Provide application’s access to Grid services
  - Provides OS independent services
- Grid enabled version of bioinformatics data management tools (e.g. DL, SRS, etc.)
  - Need to support virtual databases via Grid services
  - Grid support for commercial databases
- Bioinformatics applications “plug-in” modules
  - End user tools for a variety of domains
  - Support major existing Bio IT platforms

A Modest Proposal

- Build an Bioinformatics applications development layer on top of basic grid services
  - Think Grid enabled Matlab Toolkit for Biology
- Re-engineer bioinformatics database integration layer to target Grid services model for access
  - Virtualize access to biology databases
- Deploy a network of virtual bioinformatics “Computer Centers” leveraging existing BioGrid resources and new Grid infrastructure (e.g. TeraGrid etc.)
  - Create rich market of resources and services based on common view of BioGrid

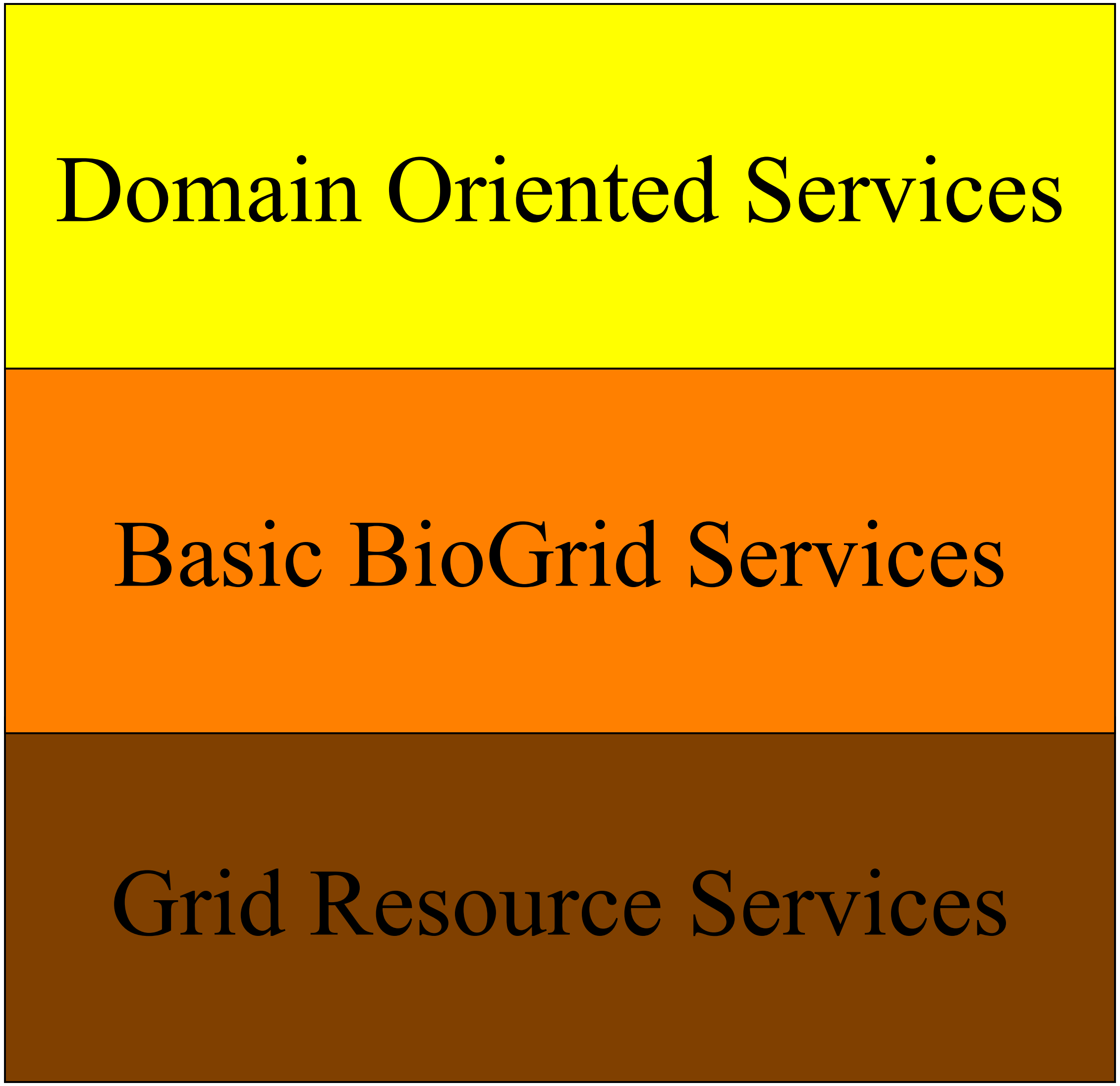
Mathematical Toolkits for Modeling Biological Systems

- "A Mathematica for molecular, cellular and systems biology"
  - Core data models and structures
  - Optimized functions
  - Scripting environment [e.g. Python, PERL, ruby, etc.]
  - Database accessors and built-in schemas
  - Simulation interfaces
  - Parallel and accelerated kernels
  - Visualization interfaces [info-vis and sci-vis]
  - Collaborative workflow and group use interfaces



Rick Stevens

BioGrid Services Model



Domain Oriented Services

- Drug Discovery
- Microbial Engineering
- Molecular Ecology
- Oncology Research

Basic BioGrid Services

- Integrated Databases
- Sequence Analysis
- Protein Interactions
- Cell Simulation

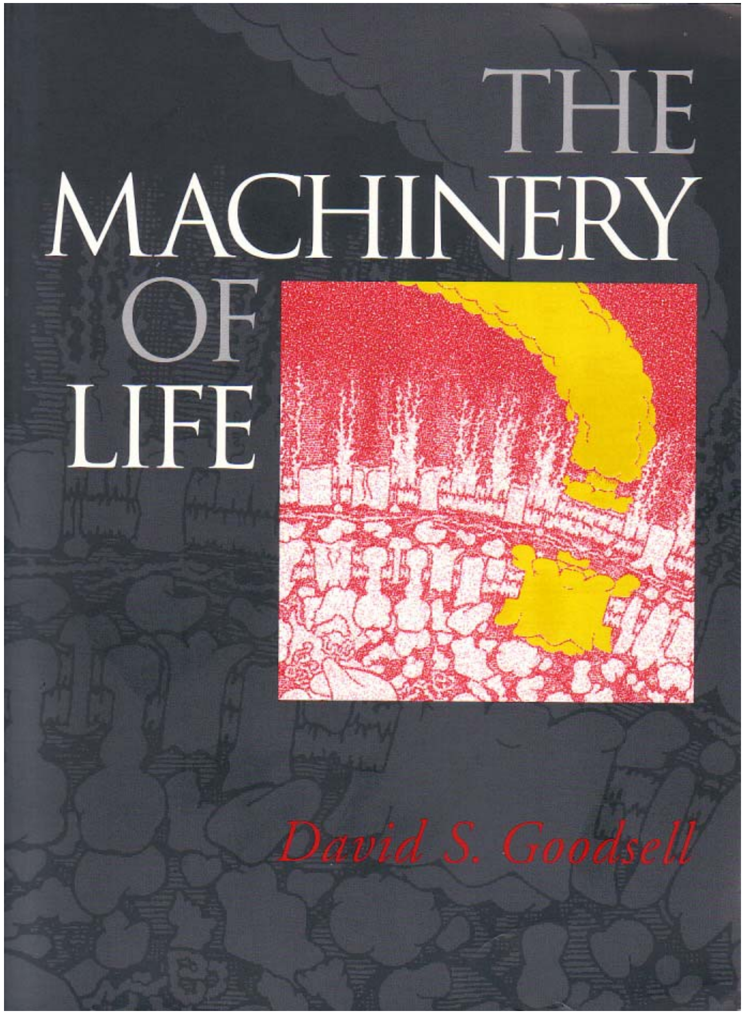
Grid Resource Services

- Compute Services
- Pipeline Services
- Data Archive Service
- Database Hosting

An International Systems Biology Grid

- A Data, Experiment and Simulation Grid Linking:
  - People [biologists, computer scientists, mathematicians, etc.]
  - Experimental systems [arrays, detectors, MS, MRI, EM, etc.]
  - Databases [data centers, curators, analysis servers]
  - Simulation Resources [supercomputers, visualization, desktops]
  - Discovery Resources [optimized search servers]
  - Education and Teaching Resources [classrooms, labs, etc.]
- Different than and more fine grain than current Grid Projects
  - More laboratory integration [small laboratory interfaces]
  - Many participants will be experimentalists [workflow, visualization]
  - More diversity of data sources and databases [integration, federation]
  - More portals to simulation environments [ASP models]

- David Goodsell
- Illustrations at 1Mx scale of many cellular domains
- Powerful use of scientific illustration
- Introduction to cellular biology and basic molecular biology



Conclusions

Acknowledgements

- DOE, NSF, ANL, UC and Microsoft support my work
- John Wooley (UCSD), Mike Colvin(LLNL/DOE), Ian Foster (ANL/UC), Jack da Silva(NCSC), Bruno Sobral(VT/VBI), Richard Gardner(InCellico) and others contributed to this talk

Thank You for Listening

Science for the 21st Century

- Relevant 20th century milestones
  - Electronic structure theory
  - Silicon based computers
  - Optical networking
  - Software engineering and open source
  - Molecular biology and genomics
  - Electron microscopy and x-ray crystallography
  - Grid computing

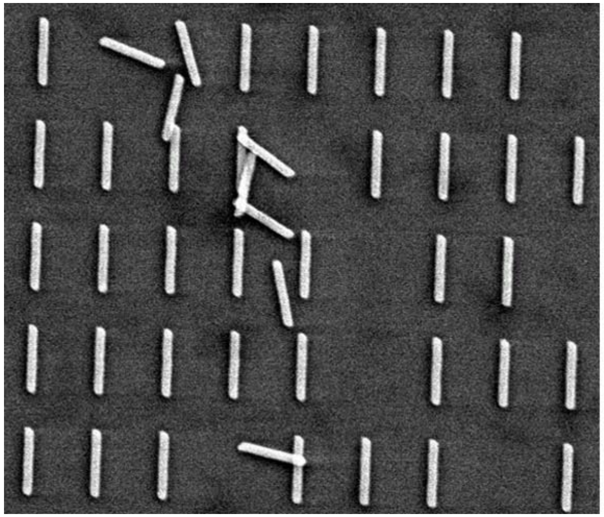
Science for the 21st Century [II]

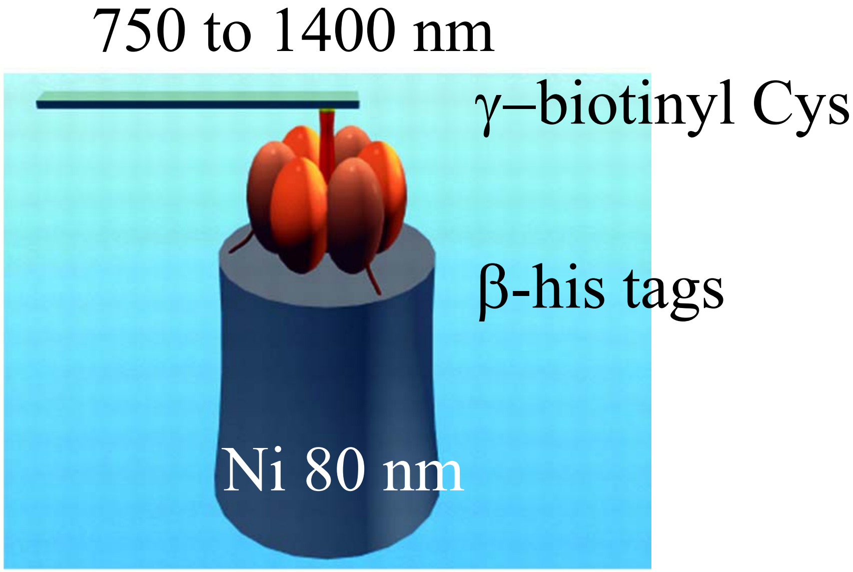
- Future science milestones
  - First synthetic model prokaryotic organism
  - Characterization of human microbial ecology
  - Global index to life on earth
  - Characterization of microbial life
  - Theory of cell evolution and organization
  - Theory of evolution of intelligence
  - First synthetic eukaryotic organism
  - Confirmation of extra-solar earthlike planets
  - Synthetic self-reproducing biomemetic nanosystem

Science for the 21st Century [III]

- The application of advanced biological thought and related technology could yield:
  - Safe and abundant food supplies
  - Sustainable and benign energy sources
  - Effective management of disease and aging
  - Novel materials and renewable industrial feedstocks
  - Advanced computational devices beyond Moore's law
  - Wide variety of molecular scale machinery
  - Self-assembly and self-reproduction technologies

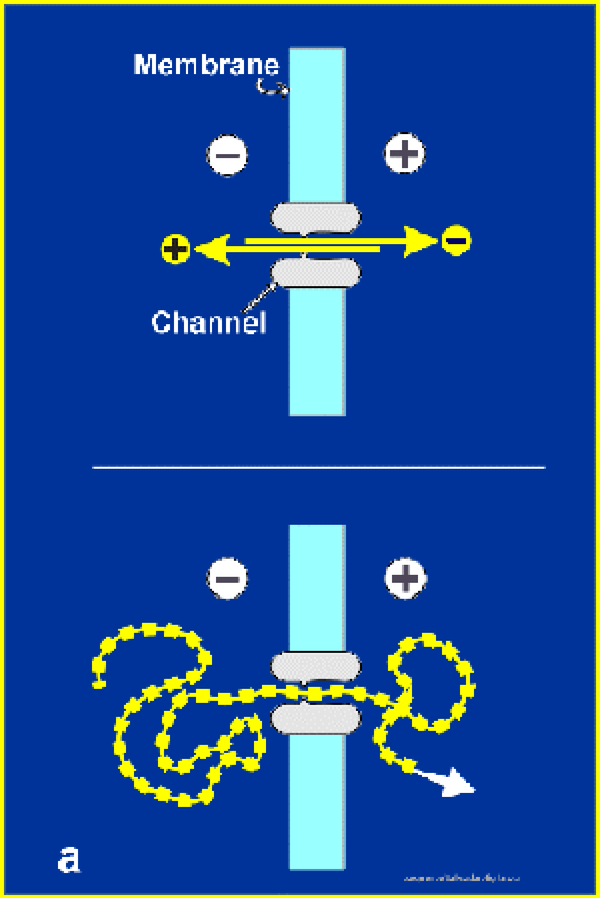
Nano-ElectroMechanical Systems (NEMS)

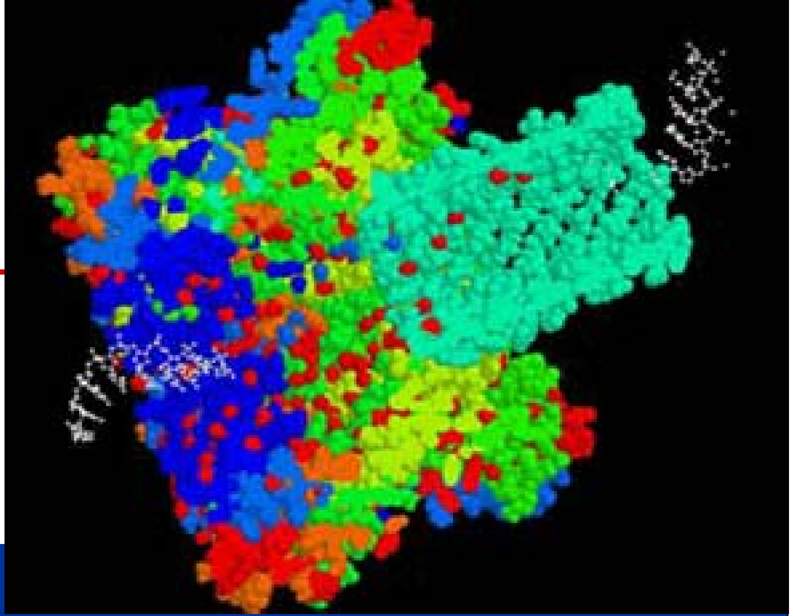


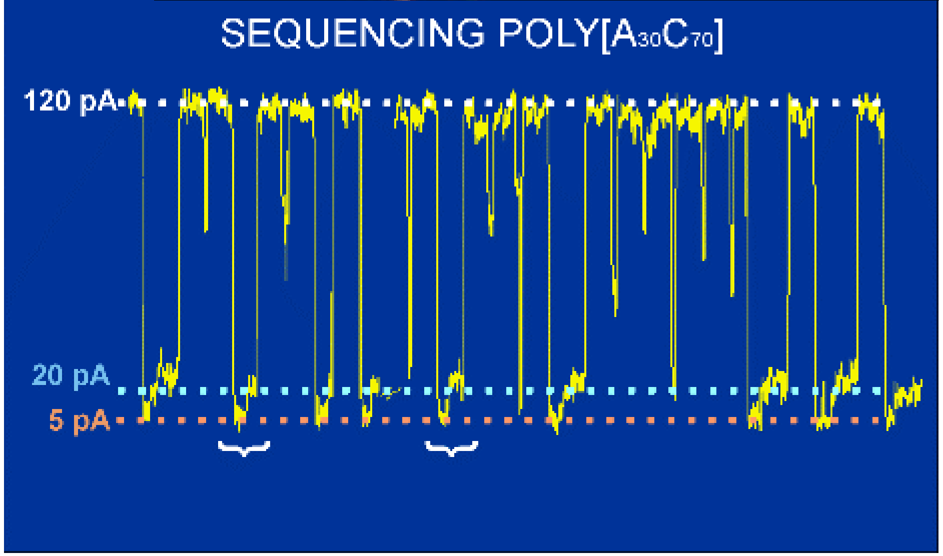


Soong et al. Science 2000; 290: 1555-1558.Powering an Inorganic Nanodevice with a Biomolecular Motor. (Pub)

Nanosensors



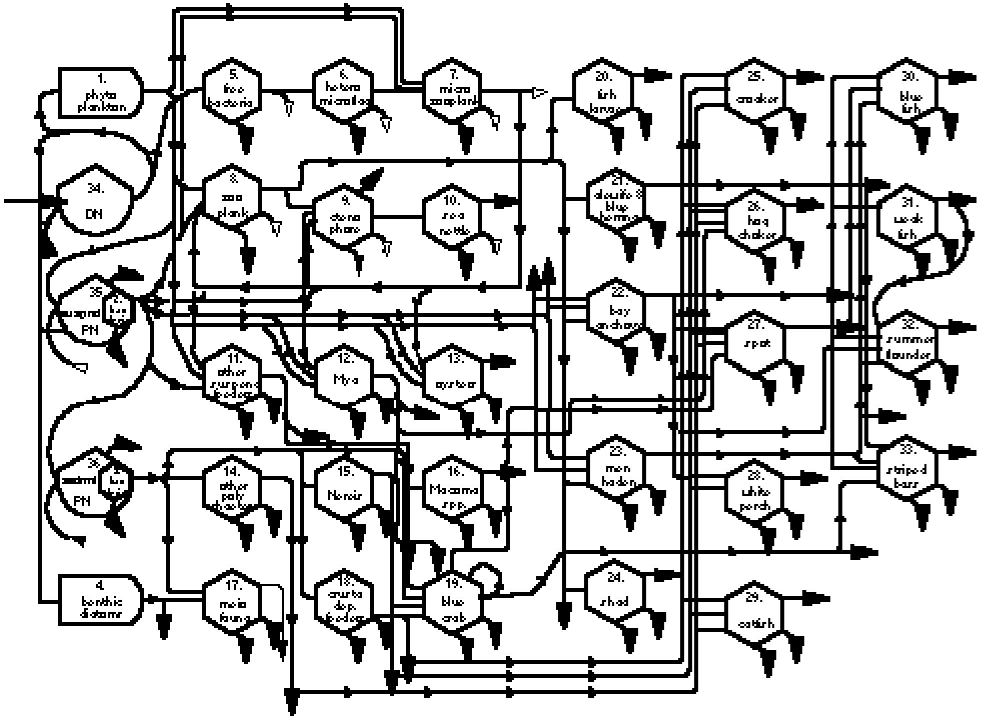




Meller, et al. (2000) "Rapid nanopore discrimination between single polynucleotide molecules." PNAS 1079-84. Akeson et al. Microsecond time-scale discrimination among polyC, polyA, and polyU as homopolymers or as segments within single RNA molecules. Biophys J 1999;77:3227-33

Agent Based Approaches to Computational Ecology

- Large-scale ecosystems models
- Individual-base models
  - Age specific behaviors
  - Goal specific behaviors
- Integration with Geochemical process cycles
- Ecosystem Network Analysis
- Microbial ecosystems



What Could the BioGrid Look Like?

- Logical access to all biological databases
  - Integrated (synthesized?) views of the data with common semantics
  - Variety of active data services
- Transparent access to analysis and modeling services
  - Over 50 commonly used tools with another 100 less common tools
  - Improved composibility of the tools in the tool chain
  - Support for canned analysis and modeling protocols
- Access to a variety of compute and storage resources
  - Services interfaces for computing platforms
  - Grid templates
- Grid enabled scripting languages
  - PERL and Python